Exploring the binding of corrin-based intermediates to enzymes of the cobalamin biosynthetic pathway

Benjamin Charles Scott

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

School of Biological and Chemical Sciences
Queen Mary University London

Required statement of originality for inclusion in research degree theses

I, Benjamin Charles Scott, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Benjamin Scott
Date:

Details of collaboration and publications:
[insert details here if applicable]

Exploring the binding of corrin-based intermediates to enzymes of the cobalamin biosynthetic pathway

Cobalamin, commonly known as vitamin B12, is the most complex small molecule created in nature. The biosynthesis is carried out in microorganisms and requires around thirty enzyme catalysed steps. Understanding its synthesis is of interest, as potentially more new analogues may be created. These would be non-toxic, water-soluble molecules, which are beginning to show promise as drug delivery systems in medical treatment.

The thesis focuses on a selection of enzymes located halfway through the aerobic cobalamin biosynthetic pathway. This includes six canonical methyltransferases as well as CobK, CobH and CobB. The structures of CobK and CobH have recently been solved and show product trapping activity. This unusual behaviour of an enzyme binding its product tightly raises two main questions. Firstly, why do these enzymes hold on to their product, a feature which may potentially help or hinder the production of cobalamin biosynthesis, and secondly, how does the subsequent enzyme in the pathway obtain its substrate?

To probe these questions a number of techniques have been explored. In Chapter 3, X-ray crystallography is used to attempt to solve the structures of CobL and CobB. Chapter 4 demonstrates the unusual product binding activity of CobK using UV-VIS spectroscopy, NMR and ITC. This work led on to the investigation of how CobL obtains its substrate which was studied using enzyme kinetics. The chapter concludes with the surprising discovery of an incorrect cobalamin intermediate binding out of turn in the pathway. This prompted an analysis of *in silico* docking data of the canonical methyltransferases in Chapter 5, to attempt to determine how the specificity of highly similar intermediates is coordinated and distinguished in enzymes with such similar architecture.

Acknowledgements

Firstly, I would like to thank my supervisor Professor Richard Pickersgill for the opportunity to complete this PhD. I am extremely grateful for his continued support and feedback on how to continue moving my project forward, as well as his wider support to attend courses to further assist in my continued learning.

I am grateful for the time my academic panel, Dr Ewan Main and Dr Maxie Roessler, put in to help review my progress and their thoughts and ideas regarding this project, particularly Ewan and his assistance in showing me how to process my kinetics data. I am also extremely grateful to Dr James Garnett for his input and assistance with crystallography, as well as Birkbeck University Structural Facilities for helping coordinate Synchrotron visits. The assistance of Dr Shuang Gu for NMR data collections and Dr Roberto Buccafusca for his Mass Spectrometry analytical services is also greatly appreciated. I am also thankful to Dr Charlotte Frankling and Miss Irene Pinzututi for their help and support regarding ITC.

I was grateful to work alongside valued laboratory colleagues Dr Nadine Younan, Dr John Wright, Dr Mark Freely, Dr Hui Zhang, Dr Ruth Rose, Miss Tania Robledo-Retana, Miss Ana Pacheco-Gutierrez and Miss Yumiko Tashiro and thank them for their continued support. Special mentions include my now good friends Dr Ismail Uddin, Dr David Bode and Mr Jack Bradley-Clarke whose guidance and encouragement was irreplaceable.

Finally, a huge thank you needs to go to friends and family and particularly my wife Becky who has been invaluable on this journey.

This work would not have been possible without the funding support of the EPSRC.

Table of Contents

Required statement of originality for inclusion in research degree theses 2
Abstract 3
Acknowledgements 4
Contents 5
List of figures 9
List of tables 11
List of equations 11
Abbreviations 12
Chapter 1: Introduction 14
1.1 What is cobalamin and how is it used in nature? 14
1.2 Applications of cobalamin-based molecules 22
1.3 How cobalamin is synthesised 23
1.4 Enzymes of interest 33
1.4.1 Canonical methyltransferases 33
1.4.2 CobK 36
1.4.3 CobL 39
1.4.4 CobH 44
1.4.5 CobB 46
1.5 Purpose of thesis 50
Chapter 2: Materials and Methods 52
2.1 Chemicals 52
2.2 Genes 52
2.3 Microbiology techniques 54
2.3.1 Media for microbial growth 54
2.3.2 Cloning and isolation of plasmid DNA 55
2.3.3 Production of recombinant proteins 57
2.3.4 Growth of cobalamin intermediate producing cells 58
2.4 Purification of recombinant proteins 59
2.4.1 Protein purification buffers 59
2.4.2 Lysis of cells 60

2.4.3 Purification of His-tagged proteins 60
2.4.4 Protein buffer exchange 60
2.4.5 Size exclusion chromatography (SEC) 61
2.4.6 SDS polyacrylamide gel electrophoresis (SDS-PAGE) 61
2.4.7 Native polyacrylamide gel electrophoresis (Native-PAGE) 61
2.4.8 Thrombin cleavage 62
2.5 Preparation of cobalamin intermediates 62
2.5.1 Lysis of cells 62
2.5.2 Purification of His-tagged protein 63
2.5.3 Protein buffer exchange 63
2.5.4 Reverse phase chromatography 64
2.5.5 Mass spectrometry (MS) 65
2.6 Spectroscopy and biochemistry techniques 65
2.6.1 UV-VIS spectroscopy 65
2.6.2 Enzyme kinetics using UV-VIS spectroscopy 66
2.6.3 Fluorescence spectroscopy 66
2.6.4 Real-time PCR thermal melts 67
2.6.5 Isothermal calorimetry (ITC) 67
2.7 Structural biology 68
2.7.1 X-ray crystallography background and methodology 68
2.7.1.1 Introduction to X-ray crystallography 68
2.7.1.2 X-ray crystallography screening and optimisation 70
2.7.1.3 Picking crystals 70
2.7.1.4 Collecting crystal diffraction data 71
2.7.1.5 Calculating the electron density map 72
2.7.1.6 Structure validation 75
2.7.2 Nuclear magnetic resonance (NMR) 75
2.7.2.1 N^{15} labelled media 75
2.7.2.2 NMR data collection 76
2.8 Bioinformatics 76
Chapter 3: Crystallography of Cob proteins 77
3.1 Introduction, aims and objectives 77
3.2 CobL crystallography 78
3.2.1 CobL protein purification 78
3.2.2 CobL crystal screening and optimisation 79
3.2.3 CobL crystal data processing 82
3.2.4 CobL co-crystallisation with CobK 84
3.3 CobB crystallography 90

3.3.1 CobB protein purification 90
3.3.2 CobB crystal screening and optimisation 93
3.3.3 CobB crystal data processing 94
3.3.4 CobB selenomethionine crystals 95
3.3.5 CobB co-crystallisation with CobH 98
3.4 Discussion and future work 103
3.4.1 CobL crystallography 103
3.4.2 CobK crystallography (CobL co-crystallisation with CobK) 105
3.4.3 CobB crystallography 107
3.4.4 CobH crystallography (CobB co-crystallisation with CobH) 109
3.4.5 Concluding remarks 112
Chapter 4: Investigating product binding and release from CobK 113
4.1 Introduction, aims and objectives 113
4.2 PC6B binding and release to CobK 116
4.2.1 PC6B purification 116
4.2.2 PC6B wavelength scans under different buffer conditions 119
4.2.3 PC6B binding titration 121
4.2.4 CobK thermal melt 123
4.2.5 NMR of CobK product binding activity 124
4.2.6 Isothermal titration calorimetry (ITC) 127
4.3 Probing protein-protein interactions between CobK and CobL 129
4.3.1 Native gel studies 129
4.3.2 Nickel column pull down 130
4.3.3 Size exclusion chromatography 133
4.3.4 Thermal melt studies of CobK and CobL 136
4.3.5 Does SAH cause PC6B release? 137
4.3.6 Enzyme kinetics 138
4.3.7 NMR of CobK with CobL^{c} 145
4.4 Promiscuous binding of cobalamin intermediates to CobK 146
4.4.1 Fluorescence studies 146
4.4.2 NMR of CobK binding HBA 149
4.4.3 CobK crystals with ligand binding 150
4.5 Discussion and future work 153
4.5.1 CobK product binding activity 153
4.5.2 Probing potential protein-protein interactions between CobK and CobL 155

4.5.3 Promiscuous binding of cobalamin intermediates to CobK 159
4.5.4 Concluding remarks 161
Chapter 5: The Cob Methyltransferases 163
5.1 Introduction, aims and objectives 163
5.2 Sequence analysis of aerobic and anaerobic methyltransferase pairs 168
5.3 Docking data of methyltransferases 174
5.3.1 Docking of PC4 (product) to CobJ (3NUT.pdb) 175
5.3.2 Docking of PC4 (substrate) and PC5 (product) to CobM (3NDC.pdb) 178
5.3.3 Docking of PC6A (substrate) to CobF (3ND1.pdb) 182
5.3.4 Docking of PC6 (substrate) to CbiE (2BB3.pdb) 184
5.3.5 Summary of docking interactions 187
5.4 Exploring the β-loop sequences and precorrin ligand decoration patterns 190
5.5 Discussion and conclusions 192
5.5.1 Sequence analyses 192
5.5.2 Docking analysis 193
5.5.3 Exploring the sequences of the β-loops and precorrin ligand decoration patterns 198
5.5.4 Concluding remarks 198
Chapter 6: Conclusions and Future Work 200
6.1 Concluding remarks of crystallography 200
6.2 Concluding remarks of CobK product trapping activity and release 201
6.3 Concluding remarks of canonical methyltransferases 203
6.4 Future work 203
References 205

List of figures

Figure 1: Alternative forms of cobalamin 15
Figure 2: An overview of the chemical structures of vitamins 17
Figure 3: T How cobalamin is absorbed 20
Figure 4: Possible tetrapyrrole derivatives from uroporphyrinogen III. 24
Figure 5: A summary of how microorganisms synthesise or salvage cobalamin 27
Figure 6: A guide on how to understand the naming convention of cobalamin synthesis.. 28
Figure 7: The biosynthetic steps of uroporphyrinogen III from 5-aminolevulinic acid. 30
Figure 8: The synthesis of cobalamin intermediates from uroporphyrinogen III to HBAD. 31
Figure 9: The final steps of cobalamin biosynthesis beyond the scope of this project 32
Figure 10: Summary of Class III methyltransferases. 35
Figure 11: A structural comparison of CobK Structures. 38
Figure 12: Proposed sequential mechanism of CobL. 40
Figure 13: How CobL^{c} monomers assemble 40
Figure 14: A structural comparison of CobL^{c} with CbiT 41
Figure 15: A structural comparison of the N- and C-termini of CbiET 41
Figure 16: A structure of the CbiET tetramer 43
Figure 17: A proposed mechanism for the formation of HBA by CobH 45
Figure 18: A proposed mechanism of CobB 48
Figure 19: Various representations of CobB 49
Figure 20: The plasmids used for Cob proteins and cobalamin intermediate production. 57
Figure 21: Example of a nickel column purification of HBA 63
Figure 22: Examples of a RP18 column purification of PC6B and HBA. 64
Figure 23 X-ray crystallography experimental pipeline and background 69
Figure 24: SDS PAGE gel and UV spectroscopy of CobL protein purification. 78
Figure 25: Thermal melt study of CobL with ligands.. 80
Figure 26: CobL crystal screening and optimisation. 81
Figure 27: L-test showing that the CobL crystals were highly twinned. 83
Figure 28: A typical CobK protein purification. 85
Figure 29: Crystals obtained from screening attempts of CobK mixed with CobL. 85
Figure 30: Electron density map of CobK disulphide bond.. 88
Figure 31: Octahedral coordination of a sodium ion binding in CobK crystal data.. 88
Figure 32: The two CobK protein chains modelled in the asymmetric unit. 89
Figure 33: A comparison of CobB protein purification with and without HBAD present. 91
Figure 34: Thermal melt studies of CobB with ligands. 92
Figure 35: Crystals obtained of CobB from optimisation trays.. 93
Figure 36: Protein purification of CobB labelled with selenomethionine. 95
Figure 37: A selection of CobB SeMet crystal optimisations. 96
Figure 38: CHOOCH fluorescence scan showing the presence of selenium in CobB crystals. 97
Figure 39: Protein purification of CobH 98
Figure 40: A selection of crystals observed in the co-crystallisation of CobB and CobH 99
Figure 41: Comparison of HBA bound CobH structure 102
Figure 42: Structural comparison of CobL^{c} and CbiET. 104
Figure 43: How XP Screen: The Protein Crystal Glue-96 solutions MiTeGen works. 105

Figure 44: Structral comparison of CobK against existing structures. 106
Figure 45: Thermal melt studies of CobH with ligand and other Cob proteins 111
Figure 46: UV spectroscopy of PC6B purification and pH change 117
Figure 47: SDS PAGE gels of new and old protocols to purify PC6B. 117
Figure 48: UV spectroscopy of PC6B RP18 purification steps. 118
Figure 49: Mass spectroscopy of PC6B purification. 118
Figure 50: UV spectroscopy of PC6B in increasing sodium chloride concentrations. 119
Figure 51: UV spectroscopy of purified free PC6B at different pH. 120
Figure 52: UV spectroscopy titrations of CobK and/or NADPH into free PC6B. 122
Figure 53: Thermal melt experiment of CobK in the presence and absence of the ligands. 123
Figure 54: Protein purification of N^{15} labelled CobK. 124
Figure 55: HSQC NMR of N^{15} labelled CobK with the ligands NADPH and PC6B 126
Figure 56: ITC titration of PC6B titrated into CobK. 128
Figure 57: Native PAGE gel of CobK and CobL 129
Figure 58: Thrombin TEV cleavage temperature trials on the proteins CobK and CobL. 131
Figure 59: Nickel column pull down experiment of thrombin treated CobK and CobL. 132
Figure 60: CobK and CobL protein purification and size exclusion chromatography 133
Figure 61: SEC pull data of CobK and CobL 135
Figure 62: Thermal melt shift of CobK and CobL together 136
Figure 63: A UV titration to explore if the addition of SAH and SAM affects product binding 138
Figure 64: A pre- and post-scan of kinetic studies of PC6B turnover. 139
Figure 65: Kinetics data for PC6B turnover in presence and absence of CobK. 142
Figure 66: A Michaelis-Menten and a Lineweaver-Burk plot of PC6B turnover. 143
Figure 67: 1D NMR of CobK with PC6B upon addition of $CobL^c$ 145
Figure 68: Wavelength scan of HBA in the presence and absence of CobK 146
Figure 69: CobK binding HBA nickel pull down and fluorescence wavelength scan 148
Figure 70: HSQC NMR of N^{15} labelled CobK in the presence of HBA. 149
Figure 71: CobK crystal optimisation produced in the presence of PC6B and NADPH 150
Figure 72: Electron density map of C5-methylated PC6B and CobK structural comparison 152
Figure 73: Pre- and Post-scan of PC6B conversion with CobL only, in CobL buffer of pH8.0 157
Figure 74: A simplified precorrin ligand docking schematic. 167
Figure 75: A sequence comparison of the Class III methyltransferases. 172
Figure 76: SAH docked in CobJ with GDP motif highlighted. 173
Figure 77: The amino acids involved in docking interactions of PC4 to CobJ. 177
Figure 78: The amino acids involved in docking interactions of PC4 and PC5 to CobM 179
Figure 79: The amino acids involved in docking interactions of PC6A to CobF. 182
Figure 80: The amino acids involved in docking interactions of PC6B to CbiE 184
Figure 81: A sequence alignment of post β-strand loop regions of the methyltransferases 191
Figure 82: A summary of precorrin ligand docking. 196

List of tables

Table 1: A summary of the aerobic and anaerobic methyltransferases..... 33
Table 2: A summary of all the cobalamin biosynthetic proteins used in this project 53
Table 3: A summary of the media components required for protein production 54
Table 4: Summary of protein purification buffers used for each Cob protein. 59
Table 5: Data collection statistics for CobL crystals 83
Table 6: Data collection and refinement statistics for CobK crystals 86
Table 7: Data collection statistics for CobB crystals 94
Table 8: Data collection statistics for CobB crystals labelled with selenomethionine 97
Table 9: Data collection and refinement statistics for CobH crystals..... 101
Table 10: A summary of the V_{max}, K_M and k_{cat} for PC6B turnover 144
Table 11: Data collection and refinement statistics for CobK with PC6B and NADPH 151
Table 12: A summary of the aerobic methyltransferases 166
Table 13: A summary of docking models for CobJ with ligand PC4 175
Table 14: A summary of docking interactions for CobJ with ligand PC4 176
Table 15: A summary of docking models for CobM with substrate PC4 178
Table 16: A summary of docking models for CobM with product PC5 178
Table 17: A summary of docking interactions for CobM with substrate PC4..... 180
Table 18: A summary of docking interactions for CobM with product PC5 181
Table 19: A summary of docking models for CobF with substrate PC5..... 183
Table 20: A summary of docking interactions for CobF with substrate PC5 183
Table 21: A summary of docking models for CbiE with product PC6B 185
Table 22: A summary of docking interactions for CbiE with product PC6B 186
Table 23: A summary of interacting amino acids in each methyltransferase 187
Table 24: A summary of which amino acids are involved in docking interactions..... 188
Table 25: A summary of the docking orientation and decoration of precorrin ligands 189
Table 26: A summary of loop interactions with respect to pyrrole ring groups 190
Table 27: A summary of all the interacting amino acid residues from section 5.2 197

List of equations

Equation 1: Beer-Lambert Law..... 65
Equation 2: Bragg's equation 72
Equation 3:Fourier transformation for calculating electron density in X-ray crystallography..... 73
Equation 4:Equation used to fit CobL activity 142
Equation 5:Equation for working out initiation rate of CobL activity 142
Equation 6: Michaelis-Menten equation 142

α	Alpha
Å	Angstrom (0.1 nanometers)
A	Acetic acid/ acetate
ALA	5-aminolevulinic acid
AMP^R	Ampicillin resistance gene
ATP	Adenine tri-phosphate
AU	Absorbance units
β	Beta
°C	Degree Celsius
cal	Calorie
Co (I)	Cobalt ion with +1 oxidation state
Co (III)	Cobalt ion with +3 oxidation state
$CobL^c$	C-terminus of CobL protein
$CobL^N$	N-terminus of CobL protein
CTP	Cytidine triphosphate
CX (where X= whole number)	Location of carbon atom in a chemical compound
Da	Daltons
DMB	dimethlybenzimidazole
DNA	Deoxyribonucleic acid
(d)UTP	(Deoxy) Uridine triphosphate
ε	Extinction coefficient
FPLC	Fast protein liquid chromatography
g	Force of gravity
HBA	Hydrogenobyrinic acid
HBAD	Hydrogenobyrinic acid-a,c-diamide
hrs/ h^{-1}	Hours/ per hour
His	Histidine amino acid
HSQC	Heteronuclear single quantum coherence
Hz	Hertz
^{111}In	Indium isotope
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal calorimetry
k	Kilo ($x10^3$)
k_{cat}	Catalytic turnover of enzyme
K_d	Dissociation constant
K_M	Michaelis constant
L	Litre
LB	Liquid broth
M- (prefix)	Mega ($x10^6$)
-M (suffix)	Molar concentration ($moldm^{-3}$)
m- (prefix)	Milli ($x10^{-3}$)
-m (suffix)	Metre
mg	Milligrams
mol	Mole
MS	Mass spectroscopy
MWCO	Molecular weight cut off
N	North
^{15}N	Nitrogen isotope with mass number 15
n	Nano ($x10^{-9}$)

(NADPH) or $NADP^+$	(Reduced) nicotinamide adenine dinucleotide phosphate
NE	North-East
NMR	Nuclear magnetic resonance
NW	North-West
OD_{600}	Optical density at 600nm
P	Propanoic acid/propanoate
PAGE	Polyacrylamide gel electrophoresis
PC2	Precorrin-2
PC3A	Precorrin-3A
PC3B	Precorrin-3B
PC4	Precorrin-4
PC5	Precorrin-5
PC6A	Precorrin-6A
PC6B	Precorrin-6B
PC7	Precorrin-7
PC8	Precorrin-8
PDB or (.pdb)	Protein Data Bank (file)
ppm	Parts per million
UV	Ultra violet
UV-VIS	UV-Visible (wavelengths)
rmsd	Root-mean-square deviation
rpm	Rotations per minute
S	South
s	Seconds
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SE	South-East
SeMet/ SelenoMet	Selenomethionine
SDS	Sodium dodecyl sulphate
S_N1	Nucleophilic substitution unimolecular reaction mechanism
SW	South-West
TCEP	Tris (2-carboxyethyl) phosphine
TEMED	Tetramethylethylenediamine
T_m	Melting temperature
2TY	2x tryptone & yeast
μ	Micro ($x10^{-6}$)
V_{max}	Maximum enzyme velocity
W	West

Chapter 1: Introduction

1.1 What is cobalamin and how is it used in nature?

Cobalamin, more commonly known as vitamin B12 is the most complex small molecule synthesised in nature, which typically requires around thirty biosynthetic enzymes[1,2]. Figure 1 shows the structure of this organometallic co-factor where cobalt is coordinated via a dimethylbenzimidazole (DMB) base, four nitrogens and an upper axial ligand, in which the chemistry of this carbon-cobalt bond is essential and unique in the reactions it facilitates[3,4]. The ability of cobalamin to coordinate different upper axial ligands allows it to carry out a diverse range of reactions[4]. Despite being such a well-studied co-factor, ongoing research still shows there is a vast amount to continue learning about its reactivity and the functions of cobalamin containing proteins[4]. Regarding its own biosynthesis these include ongoing structure-function studies of cobalamin biosynthetic enzymes and the exploration of protein-protein interactions in what appear to be the formation of multiprotein complexes, which are explored in this thesis, as well as wider applications on prokaryotic physiology and how cobalamin is regulated and involved in cobalamin dependent pathways[5].

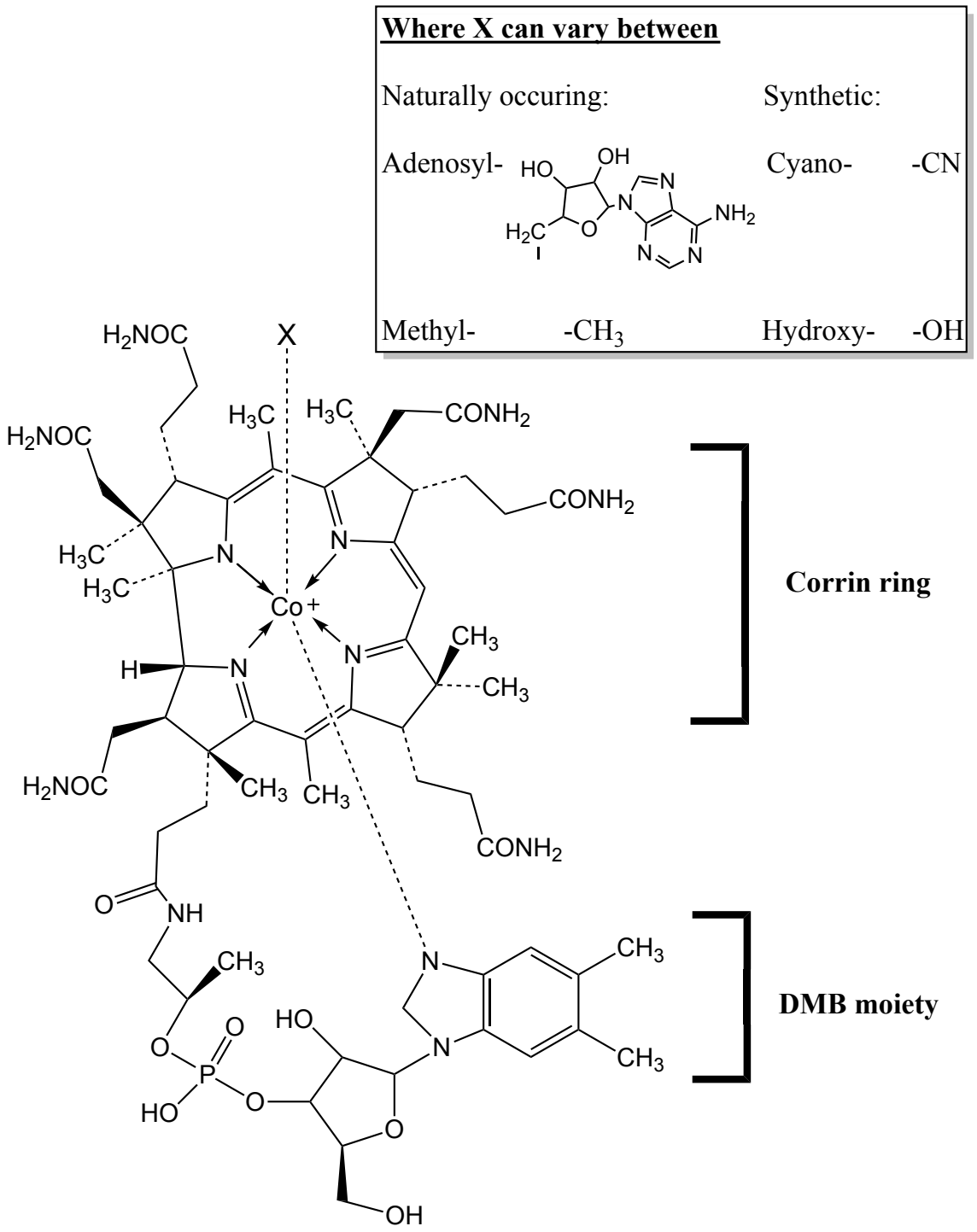


Figure 1: Shows the full structure of cobalamin where a cobalt ion is coordinated in an octahedral fashion via a dimethylbenzimidazole (DMB) base, four nitrogen atoms located in the corrin ring and an upper axial ligand labelled X. Alternative forms of cobalamin, where the upper axial ligand of cobalt, indicated by X can be substituted for a 5'-deoxyadenosyl-, methyl-, cyano- or hydroxy- chemical group. Adapted from Froese et al, 2019[6].

Vitamins and minerals are micronutrients and defined as a compound that a living organism requires but cannot synthesise itself in the required amount. Vitamins are different to minerals in that they are organic and can therefore be broken down by heat, air or acid, whereas minerals are inorganic and retain their chemical structures. The word vitamin was originally derived from the Latin '*vita',* meaning vital, and the chemical functional group '*amine',* resulting in the name *vitamine*. This therefore signified that vitamins were small organic amines found in nature which were vital for healthy (human) growth, metabolism, development and body function. However, the 'e' was later dropped when it was discovered this definition did not hold true, as not all vitamins turned out to be amines, resulting in vitamins being defined by their solubility and importance to living organisms[7]. The continued reassignment of organic molecules over time additionally led to the originally defined vitamins F-J being either reassigned or removed as they were no longer classified as vitamins. This explains why the numerical and alphabetical sequences no longer hold true in some classes of vitamins[7]. For example, vitamin B7, biotin, was originally called vitamin H. Figure 2 shows an overview of all the vitamins and their solubilities.

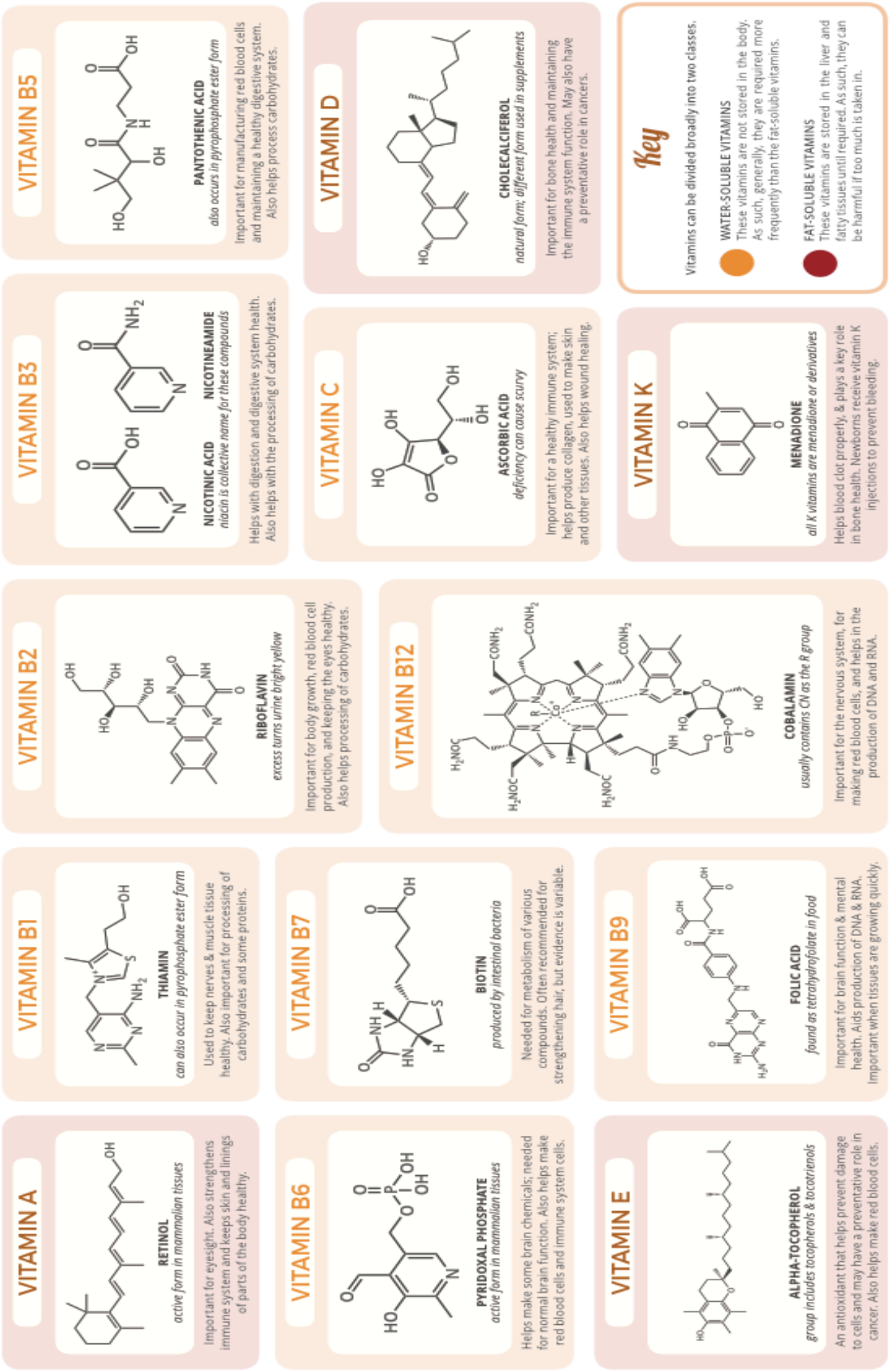


Figure 2: An overview of the chemical structures of vitamins and their respective solubilities in water or fats. Image taken from www.compoundchem.com

The in-depth study of vitamins has always been difficult, due to their involvement in enzyme catalysis as co-factors. This has meant organisms only require them to be present in low quantities resulting in the use of classical techniques such as enzyme assays being difficult to use when monitoring their activity. As a consequence the study of intermediates involved in vitamin synthesis has proved difficult to investigate, as they are typically labile or controlled by negative feedback loops[8]. It is therefore not surprising in this case that the study of cobalamin has proved to be a challenging one, involving decades of research and it has been interesting to observe in the literature how advances and improvements in science have continued to deepen our understanding of cobalamin biosynthesis and its importance in nature[9,10,11]. In fact, the research of cobalamin has resulted in two Nobel prizes being awarded. The first Nobel prize was awarded to Minot and Murphy in 1934, for discovering that vitamin B12 cured pernicious anaemia[12], and the second Nobel prize was awarded in 1964 to Dorothy Hodgkin for solving the structure of vitamin B12 using X-ray crystallography[13].

When investigating cobalamin metabolism and biosynthesis in nature there are different categories to consider; (i) organisms which do not synthesise or require cobalamin; (ii) organisms which synthesise cobalamin and require it and; (iii) organisms which require it but cannot synthesise it. There is no evidence, including from genome sequencing, to suggest that eukaryotes can synthesise cobalamin[14]. This means that humans fall into the third category, while other eukaryotes such as plants and fungi fall into the first category[14,15]. Interestingly about half of all algae also require cobalamin and differences in cobalamin availability and cobalamin independent metabolic pathways are presumed to have contributed to heterogenous occurrences of cobalamin-dependent enzymes in algae[16].

In fact, cobalamin dependence has resulted in varying symbiotic relationships. These include bacteria colonising the extracellular mucous layer of algae where photosynthetic products and bacterial cobalamin are exchanged [16]. *Caenorhabditis elegans* fed with cobalamin producing bacteria

resulted in accelerated development and reduced mortality due to a change in gene expression [17,18], while legume plants contain cobalamin dependant bacteria and when grown under cobalamin deficient conditions resulted in an inhibition of growth, due to a lack of nitrogen fixation [19].

As humans cannot synthesise cobalamin it is therefore classified as an essential vitamin in our diet. Despite having a gut flora which contains cobalamin producing microbes Degan et al, 2014 argue that these would be unlikely to provide sufficient supplies of cobalamin exceeding their own demands[16]. This therefore means our cobalamin intake is obtained from our diet and typically includes animal products such as red meat, fish and dairy produce [20]. In a similar fashion to prokaryotes our bodies have developed their own specific means to absorb and salvage cobalamin from our diets. This is summarised concisely in Figure 3. As humans ingest food containing cobalamin, which is typically present bound to cobalamin binding proteins, it is released when these are digested in the stomach. The three essential proteins involved in cobalamin binding and subsequent absorption are haptocorrin, intrinsic factor and transcobalamin. Haptocorrin is found in many bodily fluids such as saliva and bile where it serves to protect cobalamin during passage through the acidic environment in the stomach. This allows for the succeeding transfer of cobalamin to intrinsic factor, which can be absorbed via receptor mediated endocytosis in the ileum. As cobalamin enters the blood stream it binds to transcobalamin which is responsible for delivering cobalamin to cells, again via receptor mediated endocytosis, with the majority of cobalamin stored in the liver[21].

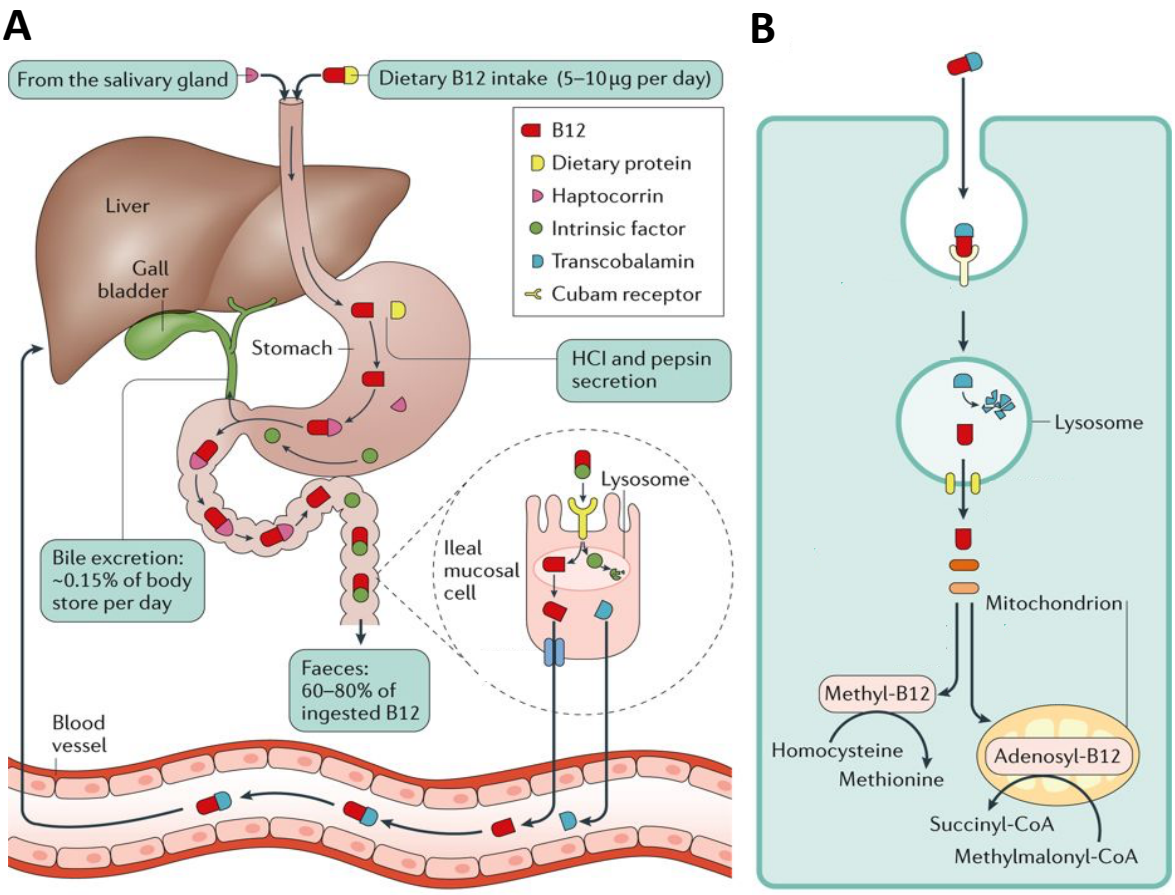


Figure 3: Taken from Green et al, 2017 [21]. A) Shows how cobalamin is absorbed into the bloodsteam via the cobalamin binding proteins haptocorrin, intrinsic factor and transcobalamin. B) Shows how cobalamin is taken up by cells, from the bloodstram, via receptor mediated endocytosis and the essential reactions methycobalamin and adenosylcobalamin are involved in for the production of methionine and succinyl-CoA respectively.

Cobalamin deficiency can arise in a number of ways including inadequate intake, inadequate bioavailability or malabsorption[21]. Green et al, 2017 provides a comprehensive summary of the different forms and reasons that cobalamin deficiency may arise due to disruption of cobalamin uptake and transport through the body[21]. Pernicious anaemia is the most well known outcome of cobalamin deficiency and is more typically seen in those living in affluent areas, where it develops as an autoimmune response, caused by a lack of intrinsic factor. Those living in low income countries are more likely to have a low intake of cobalamin due to a lack of cobalamin rich food of animal origin. Surprisingly a study by Yajnik et al, 2006 found that those living in slums with poorer hygiene in fact had a higher cobalamin intake than urban dwellers due to a higher exposure and ingestion of cobalamin containing microorganisms[22].

Humans are able to use both adenosylcobalamin and methylcobalamin for vital cellular metabolism as summarised in Figure 3B. These include DNA, amino acid and fatty acid synthesis as well as in mitochondrial metabolism[23,24]. Methylcobalamin is essential for the methylation of the amino acid cysteine into methionine. It is also involved in the formation of S-adenosylmethionine (SAM) from S-adenosylhomocysteine (SAH) which is an essential methyl donor for numerous cellular reactions including the production of the nucleotide thymine, neurotransmitters, proteins as well as epigenetic reactions such as the methylation of DNA, histones and other gene expression regulators[21,24,25]. Therefore, a lack of methylcobalamin can result in high levels of SAH, which leads to cellular stress and apoptosis. If thymine synthesis is prevented this leads to the incorrect incorporation of dUTP into DNA which has a large impact on areas of the body with high turnover such as the haematopoietic systems in bone marrow, resulting in anaemia from abnormally large blood cells. Adenosylcobalamin is involved in the formation of succinyl-CoA, an essential component of respiration in the mitochondria[21].

Overall cobalamin can be summarised as essential for the following three classes of reactions in nature:

1) **Methyl transfer reactions** occur via methylcobalamin and typically are carried out by a methyl-cation transfer. This involves a heterolytic cleavage of the carbon-cobalt bond, where the cobalt shuffles between Co(III) and Co(I)[4]. Examples include; methionine synthase and O-demethylases[26–28].
2) **Isomerase reactions** occur via adenosylcobalamin. This is by the formation of a highly reactive transient adenosyl radical by a homolytic cleavage of the carbon-cobalt bond[4]. The most notable example would be methylmalonyl-CoA mutase[29,30].
3) **Reduction reactions** via dehalogenation are unique in nature and have allowed microorganisms to live in unusual locations, as it allows them to use organohalides as the final electron acceptor in respiration[4,31]. This class of reaction has led to further study of catalysis as a means to minimise waste and disposal methods of organic halides which have

negative impacts on the environment and health[4,32]. Some examples include the enzymes CprA and haloalkane dehalogenases, which are involved in reductive dehalogenation of halogenated aromatic and aromatic hydrocarbons in microbes[33–35].

1.2 Applications of cobalamin-based molecules

A very recent study by Lawrence et al, 2018 has described how they created fluorescent analogues of cobalamin which they were then able to detect in bacteria, worms and plants[36]. This was achieved by replacing a SAM molecule with an allyl analogue in the biosynthetic pathway, upon which a fluorescent tag could be attached. Through careful tailoring of subsequent biosynthetic enzymes after this step, including site directed mutagenesis to accommodate the allyl intermediate, an incomplete cobalamin analogue was produced. Bacteria could salvage this corrinoid compound and successfully convert it into fluorescent cobalamin for metabolism. Humans are unable to salvage incomplete corrinoids therefore making this a promising application for observing and targeting cobalamin dependent infections such as Tuberculosis. Another surprising finding of this study was demonstrating the uptake of cobalamin in plants, despite them having no dependence for it, showing that nutrient-enriched plants could be grown to help tackle cobalamin deficiency in some countries.

Rapidly growing cells, such as cancer cells, have an increased demand for nutrients and vitamins[37]. This makes cobalamin an attractive therapeutic target as it is water soluble with no known toxicity and as it is an essential vitamin, it is therefore unlikely that mutations involved in uptake will arise. This has allowed radiolabelled ^{57}Co and ^{111}In forms of cobalamin to be used to image and locate tumours[37–39]. Waibel et al, 2008 went one step further to edit the radiolabelled cobalamin in such a way to minimise undesirable radio-accumulation in the kidneys. This was achieved by interrupting how it interacted specifically with transcobalamin II, while maintaining normal interaction with transcobalamin I, typically present in higher amounts in tumours[37,40].

Lodowski et al, 2018 described the use of antivitamins, which are a class of cobalamin analogues gaining recent interest in several research areas, as their photo-properties offer opportunities to explore the chemistry and metabolism of vitamin B12 alongside further applications[41]. Examples of these include photocatalysis of iridium(III) B12 complexes for reductive dehalogenation reactions[32], photolytic cleavage to permit drug-release of cob(III)alamins and ferric porphyrins[42,43] and photopolymerisation of B-12-mediated hydrogels[44].

By studying the biosynthetic enzymes of cobalamin, not only a greater understanding of its synthesis can be gained, but also the biosynthetic pathway could be altered, using mutations, to allow the creation of new molecules based on cobalamin. The advantage of such molecules is that they would be non-toxic and water soluble. Such molecules have been shown to have applications in therapeutics such as trojan horse drug delivery systems and imaging or various light inducing applications, all offering exciting potentials in many aspects of society.

1.3 How cobalamin is synthesised

Cobalamin is unique to the B series of vitamins as it is in a family of molecules known as tetrapyrroles, in which four pyrrole groups are joined together to form a conjugated carbon ring. The most well-known examples of these molecules are haem and chlorophyll. These molecules contain a planar carbon ring which is able to chelate a metal ion in the centre, via the nitrogen atom from each of the four pyrrole groups which forms the ring. Typically, this conjugated tetrapyrrole ring contains twenty carbon atoms, except in the case of cobalamin where it has a contracted ring of nineteen carbon atoms. Different tetrapyrroles chelate different metals and in the case of cobalamin the metal is cobalt, which can change its oxidation state between +1 to +3 and is essential for the reactions it helps catalyse. Tetrapyrroles are all derived from a precursor known as uroporphyrinogen III, highlighted in Figure 4, which serves as a key branch point in the biosynthesis of these various tetrapyrrole molecules.

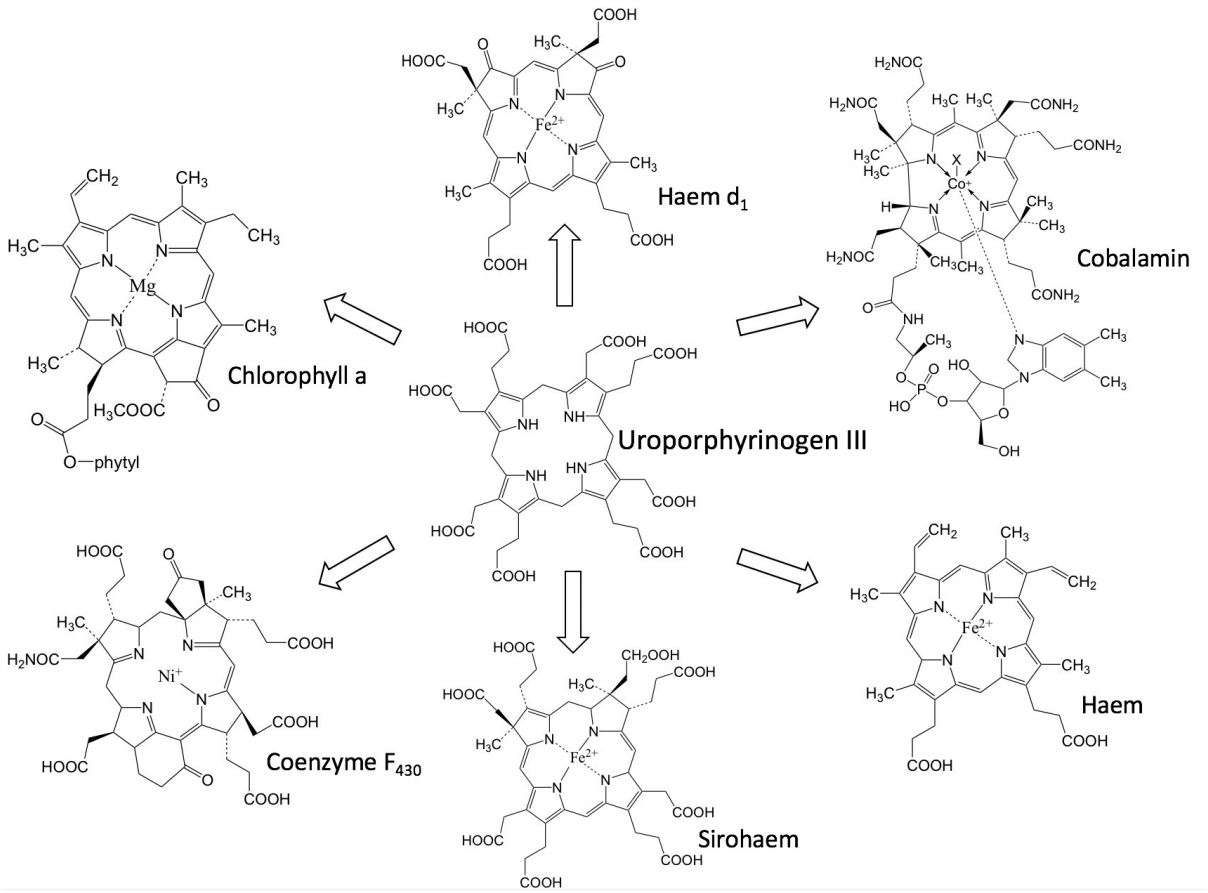
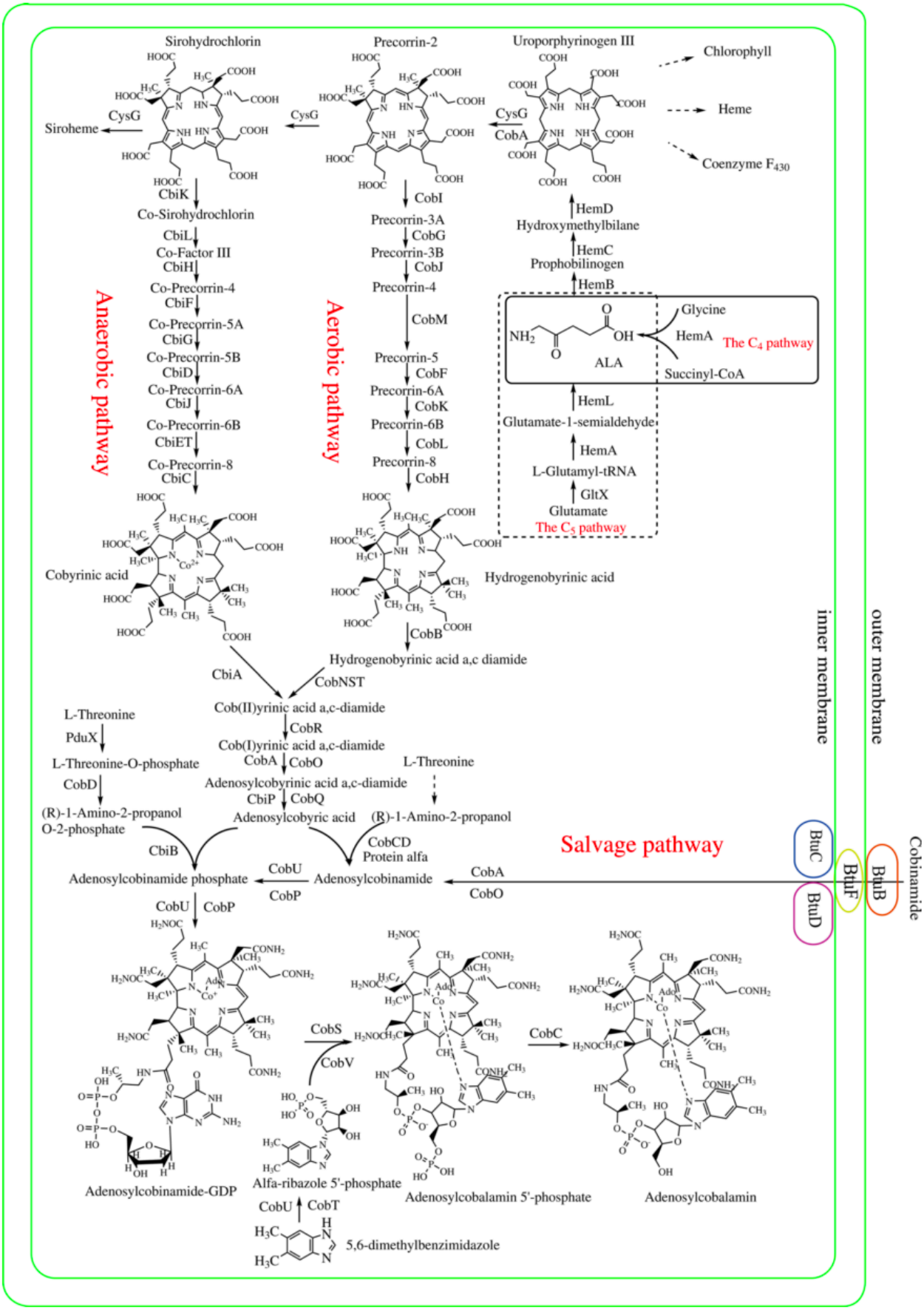


Figure 4: Tetrapyrrole derivatives which can be formed from the precursor uroporphyrinogen III. Figure adapted from Raux, et al, 2000 [1].

The structure of cobalamin can be split into three parts based on how the cobalt is coordinated via its octahedral geometry. The lower axial ligand is coordinated via a nitrogen on a ribonucleotide dimethylbenzimidazole group which loops down from the tetrapyrrole ring at carbon number 17 (C17). The four planar co-ordinations result from the afore mentioned pyrrole nitrogen atoms in the tetrapyrrole ring. Finally, the upper axial coordination varies depending on the type of cobalamin. Two naturally occurring biological forms exist where the upper coordination is either via a methyl- or adenosyl- group, named methylcobalamin or adenosylcobalamin respectively[15]. A synthetic cyanocobalamin and hydroxycobalamin also exists where the upper axial ligands are cyanide or hydroxy- groups respectively, in which the cyano- form is used in vitamin supplements. The structure of cobalamin is summarised in Figure 1 where the X can be substituted for the various upper axial ligands.

The complete *de novo* biosynthesis of cobalamin typically involves around thirty enzymes and can be carried out either anaerobically or aerobically[1,2]. The enzymes involved in aerobic synthesis are typically coined *Cob* proteins, while those involved in the anaerobic pathway are coined *Cbi*. Interestingly these two pathways are remarkably similar with the stepwise chemical modifications occurring in the same order, due to genetically and structurally similar proteins[45]. The only major difference between the two pathways, excluding oxygen usage, is that the cobalt is inserted into the corrin ring early in anaerobic synthesis and later during the aerobic pathway. The stepwise modifications of cobalamin biosynthesis are summarised in Figure 5. This allowed a wider study of how the cobalamin biosynthetic pathway proceeds and in some cases has provided excellent structural homology for proteins when one side of the pathway has been unknown. Interestingly Thermes et al, 1996 attempted using cobalt containing intermediates from the anaerobic pathway on the aerobic enzymes but found cobalamin biosynthesis was not possible by interchanging these pathways[46]. Alternatively a more cost-effective method for bacteria and archaea is to salvage cobalamin from corrinoid molecules in their local environment and transform these into adenosylcobalamin[47]. This is mediated by the Btu proteins, where BtuB is an essential transport protein and controlled by a cobalamin riboswitch[16,48,49].

A


Sirohydrochlorin
Precorrin-2
Uroporphyrinogen III
Chlorophyll
Heme
Coenzyme F430
CysG
Siroheme
CysG
CysG
CobA
CbiK
Co-Sirohydrochlorin
CobI
Precorrin-3A
HemD
Hydroxymethylbilane
CbiL
Co-Factor III
CobG
Precorrin-3B
HemC
Prophobilinogen
CbiH
Co-Precorrin-4
CobJ
Precorrin-4
HemB
CbiF
Co-Precorrin-5A
CobM
Glycine
HemA
The C4 pathway
ALA
Succinyl-CoA
CbiG
Co-Precorrin-5B
Precorrin-5
CobF
CbiD
Co-Precorrin-6A
Precorrin-6A
CobK
HemL
Glutamate-1-semialdehyde
CbiJ
Co-Precorrin-6B
Precorrin-6B
CobL
HemA
L-Glutamyl-tRNA
CbiET
Co-Precorrin-8
Precorrin-8
CobH
GltX
Glutamate
The C5 pathway
CbiC
Anaerobic pathway
Aerobic pathway
Cobyrinic acid
Hydrogenobyrinic acid
CobB
Hydrogenobyrinic acid a,c diamide
CbiA
CobNST
Cob(II)yrinic acid a,c-diamide
L-Threonine
CobR
PduX
Cob(I)yrinic acid a,c-diamide
L-Threonine-O-phosphate
CobA
CobO
L-Threonine
CobD
Adenosylcobyrinic acid a,c-diamide
CbiP
CobQ
(R)-1-Amino-2-propanol
O-2-phosphate
Adenosylcobyric acid
(R)-1-Amino-2-propanol
CbiB
CobCD
Protein alfa
Salvage pathway
inner membrane
outer membrane
BtuC
BtuF
BtuB
BtuD
Cobinamide
Adenosylcobinamide phosphate
CobU
CobP
Adenosylcobinamide
CobA
CobO
CobU
CobP
CobS
CobV
CobC
Adenosylcobinamide-GDP
Alfa-ribazole 5'-phosphate
Adenosylcobalamin 5'-phosphate
Adenosylcobalamin
CobU
CobT
5,6-dimethylbenzimidazole

B

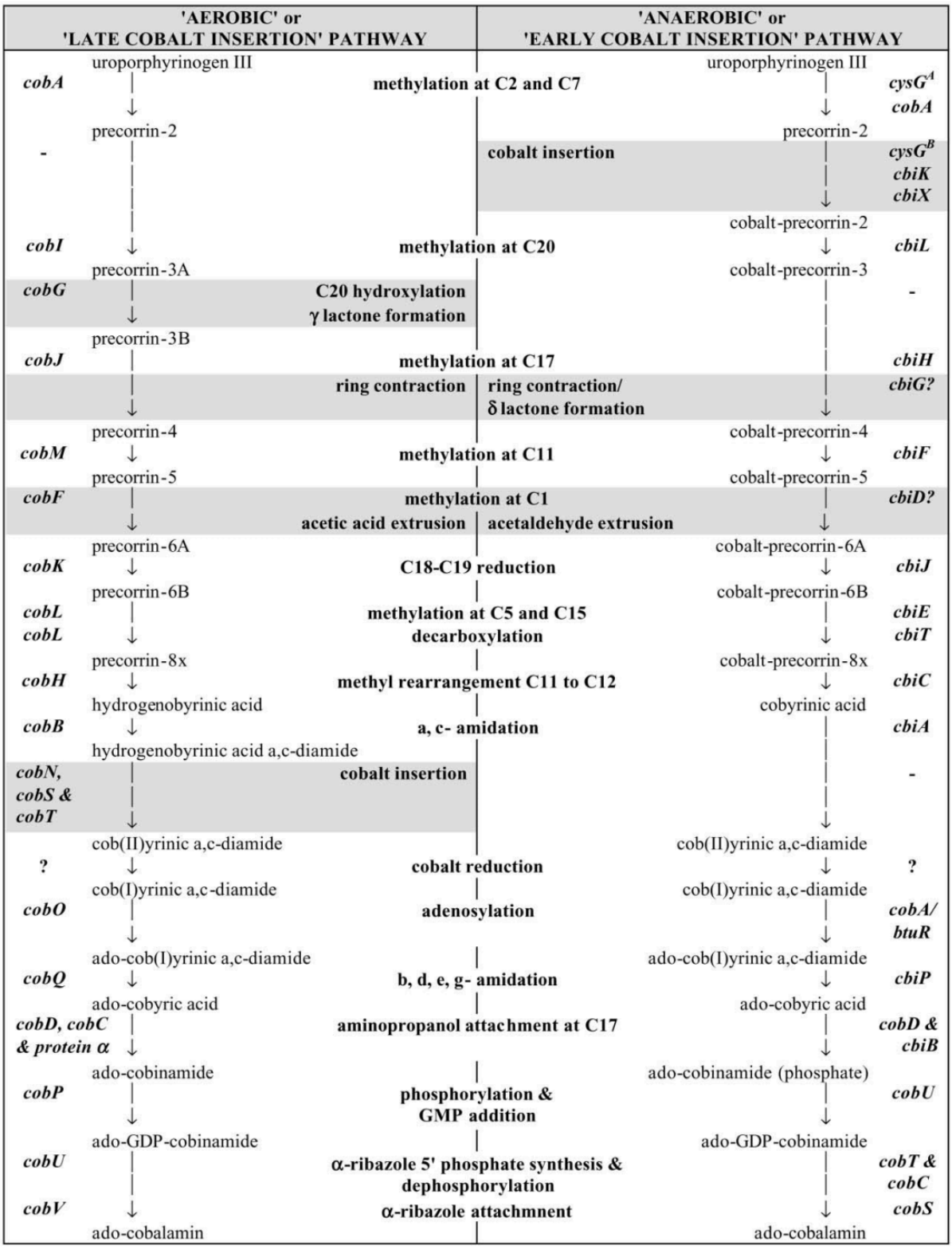


Figure 5: A) A comprehensive summary of how microorganisms synthesise or salvage cobalamin. The branch points for other tetrapyrrole intermediates are also shown as well as the alternative enzymes involved in the aerobic or anaerobic pathways. Figure taken from Fang et al, 2017[47]
B) Highlights more closely the similarities and differences of the aerobic and anaerobic synthesis of cobalamin, taken from Warren et al, 2002[2]. The chemical changes are summarised down the middle while the substrates and gene names flow down the outer edges of the two columns. The sections highlighted in grey signify major genetic differences between the two pathways.

To understand the naming conventions which describe the enzymatic reaction steps of the cobalamin biosynthetic pathway, it is important to know how the tetrapyrrole ring is labelled. The four pyrrole rings are assigned uppercase letters from A-D, which are assigned in a clockwise fashion as shown in Figure 6A. The carbon atoms of the tetrapyrrole ring are subsequently numbered from one to twenty where the numbering begins and ends at the first pyrrole group A, Figure 6B. Finally, the outer ethanoic and propanoic sidechains are labelled using lowercase letters a-g in a clockwise fashion, again beginning with pyrrole A, Figure 6C, where one of the side chains is ignored between d and e, as it is removed during cobalamin biosynthesis.

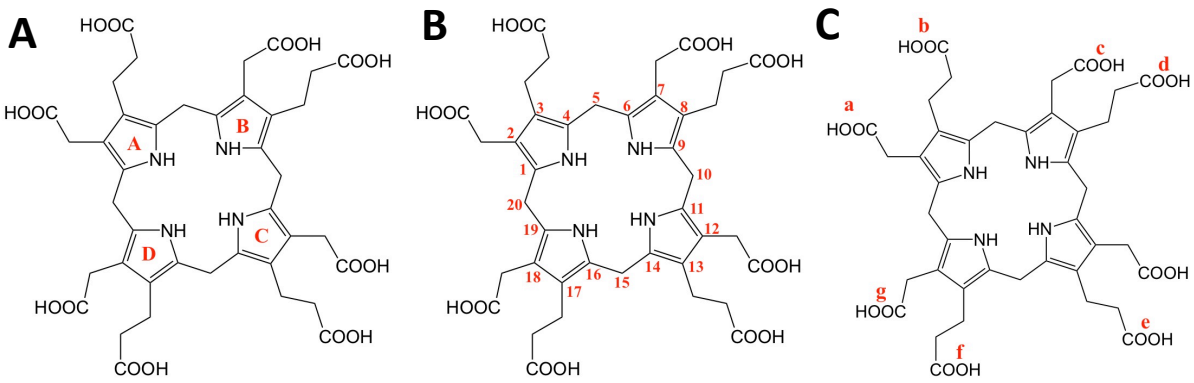


Figure 6: A guide on how to understand the naming convention used when describing the intermediates of cobalamin synthesis. A) Shows how the pyrrole rings are labelled in a clockwise fashion. B) Shows how the carbon atoms are named in a clockwise fashion. C) Shows how the carboxylic acid side chains are labelled in a clockwise fashion.

Cobalamin contains a nineteen membered heterocyclic carbon ring known as a corrin ring, Figure 1. This distinguishes corrin based molecules from porphyrins which are twenty membered heterocyclic carbon rings, found in all other tetrapyrrole based molecules[1]. The name of the intermediate uroporphyrinogen III contains '*uro-*' which comes from the Greek meaning '*combined*' indicating a combined porphyrin structure derived from pyrrole groups. The roman numeral 'III' identifies the third isomeric form, based on how the carboxyl side chains are arranged, with the third form found exclusively in nature[11,50]. Subsequently many of the cobalamin intermediates are called precorrin-x. This refers to the intermediates of corrin ring biosynthesis between uroporphyrinogen

III and cobyrinic acid, with the *x* indicating the number of methyl groups which have been attached[5]. Finally, in the aerobic pathway a name change to hydrogenobyrinic acid occurs after precorrin-8, because it is the hydrogen (cobalt-free) form of cobyrinic acid from the equivalent stage of the anaerobic pathway, signifying no further methylation reactions take place[51].

For the purpose of this project the work involved on biosynthesis focuses only on the aerobic pathway. The biosynthesis of cobalamin can be broken down into three sections based on the enzymes involved in relation to the project; (i) the synthesis of the precursor uroporphyrinogen III via Hem proteins[52]; (ii) the Cob enzymes used in the laboratory for the synthesis of cobalamin intermediates up to hydrogenobyrinic acid-a,c-diamide and (iii) the remaining Cob proteins involved in cobalamin synthesis (which is beyond the scope of this project).

The initial genes involved in the synthesis of uroporphyrinogen III produce the Hem proteins. The initial precursor for any tetrapyrrole is 5-aminolevulinic acid (ALA). There are two separate pathways, which vary in different organisms, coined the C4 and C5 pathway based on the number of carbons present in the substrate for ALA synthesis. Each pathway contains a HemA protein and these are briefly summarised in the top right corner of Figure 5A. Once ALA has been produced the protein HemB subsequently combines two molecules of ALA to form a pyrrole molecule. The remaining Hem proteins HemC and HemD then combine the four pyrrole groups together and close the tetrapyrrole ring to form uroporphyrinogen III, as shown in Figure 7. The precursor uroporphyrinogen III acts as a major branch point for other tetrapyrrole derivatives as shown in the top right corner Figure 5A. Despite *Escherichia coli (E. coli)* being unable to produce cobalamin, it could be used in the laboratory for this project by hijacking the existing Hem genes involved in siroheme synthesis and subsequently introducing the Cob genes on a plasmid vector, which would produce the enzyme to act upon uroporphyrinogen III.

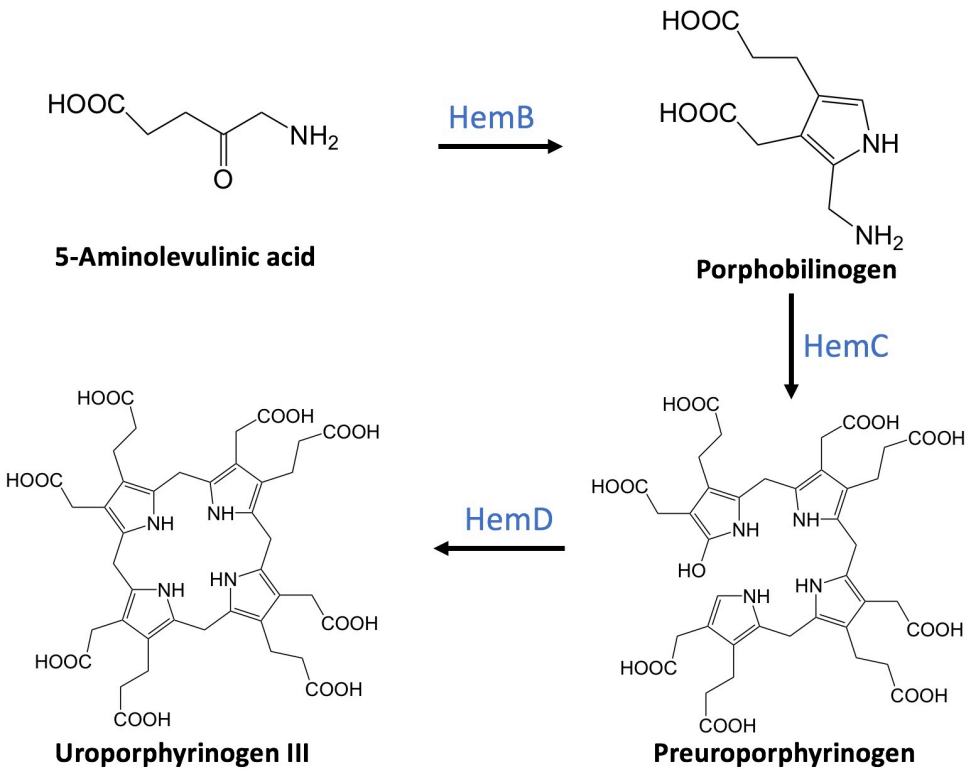


Figure 7: The initial steps involved in the formation of cobalamin in which uroporphyrinogen III, the precursor for tetrapyrrole derivatives, is formed from 5-aminolevulinic acid via the enzymes HemB, HemC and HemD.

The second part of the cobalamin synthesis focuses on the initial Cob proteins and covers the main enzymes which are the focus of this project. The stepwise enzymatic reactions of different cobalamin intermediates are shown in Figure 8. These include contraction of the tetrapyrrole carbon ring, via the removal of carbon-20, and a number of smaller modifications, including several methylations via S-adenosylmethionine (SAM). These eight methylation reactions to the corrin ring are carried out by six canonical methyltransferases and one non-canonical methyltransferase, upon which the precorrin intermediates are named and identified[50]. For example, precorrin-4 indicates the addition of four methyl-groups, while precorrin-8 indicates eight methyl-group additions. In some cases, the precorrin intermediate ends in the letter A or B, as in the case of precorrin-3 and precorrin-6, which indicates no additional methyl-groups have been added but rather signify an alternative reaction has taken place.

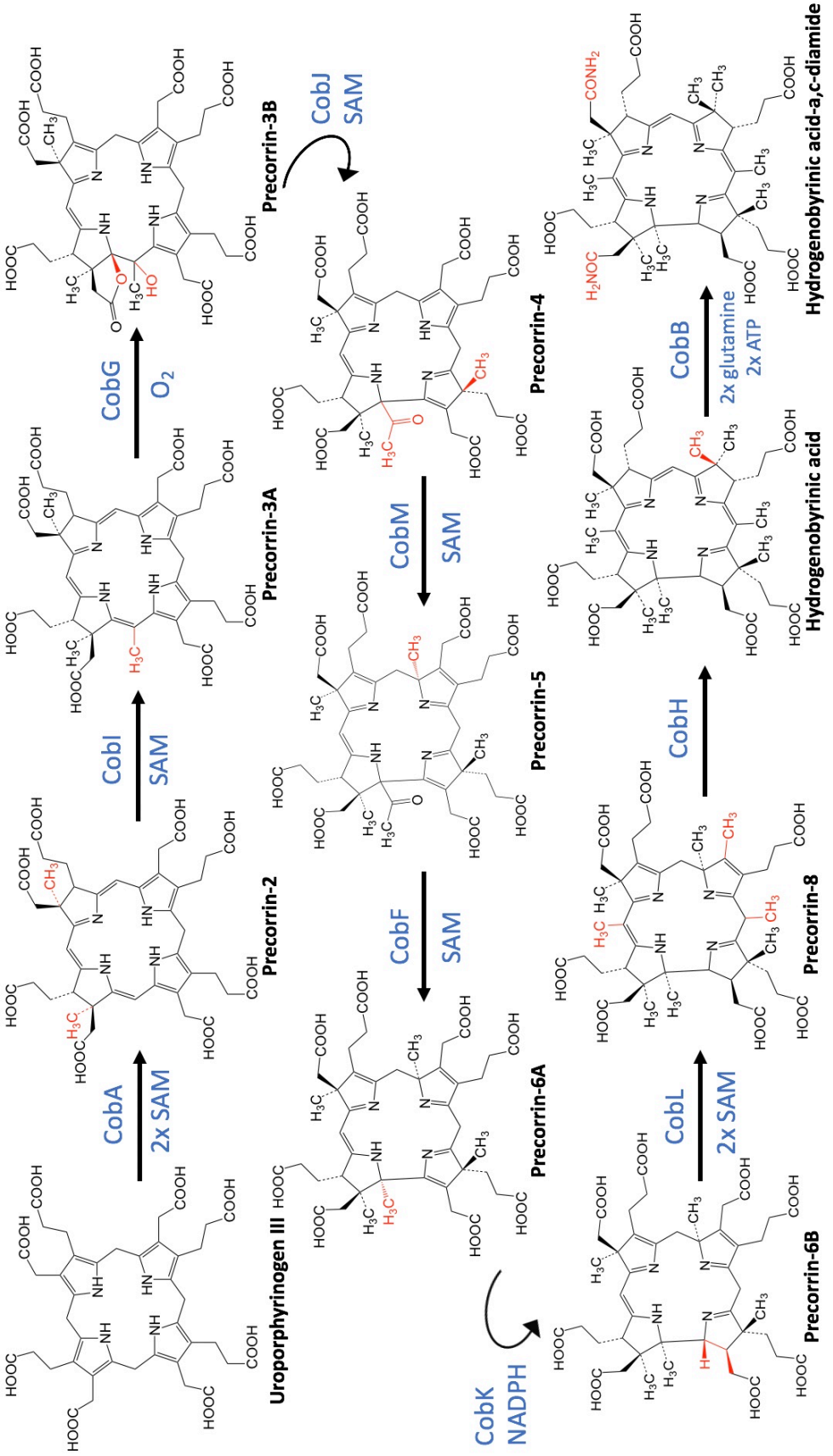


Figure 8: The step by step synthesis of cobalamin intermediates from uroporphyrinogen III to hydrogenobyrinic acid-a,c-diamide. Figure adapted from Deery et al[53].

The final section of the cobalamin biosynthetic pathway includes the insertion of cobalt into the corrin ring and the synthesis and addition of the lower and upper axial coordinated ligands. This involves connecting the ribonucleotide dimethylbenzimidazole group via a linker arm on C17 and attaching a methyl- or adenosyl- group above depending on the final form of cobalamin produced, Figure 9.

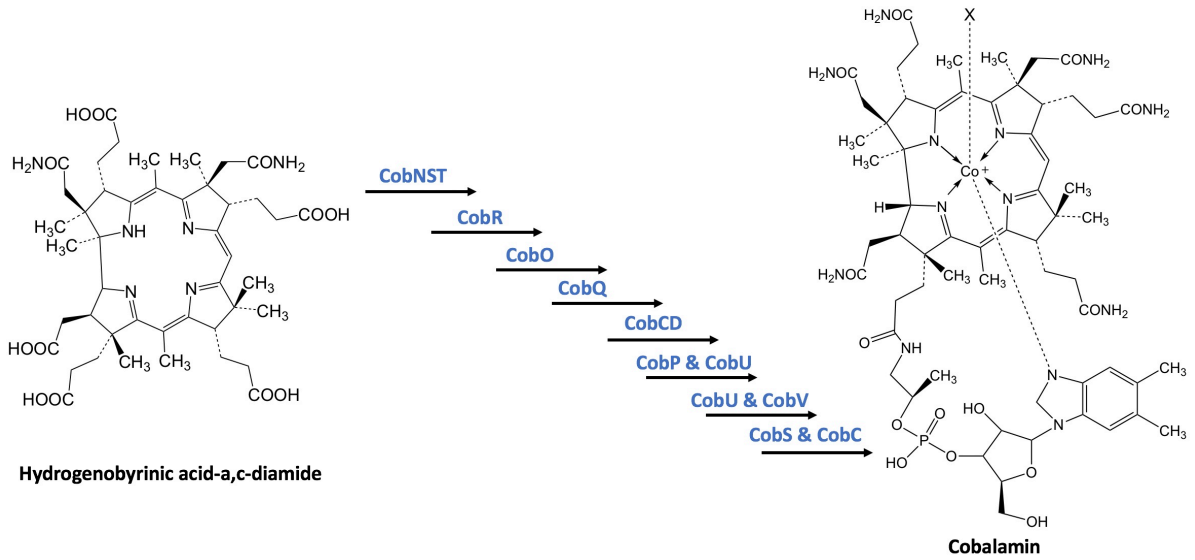


Figure 9: The final Cob enzymes inovoled in the biosynthesis of cobalamin biosynthesis that follows on from hydrogenobyrinic acid-a,c-diamide. These enzymes include the insertion of cobalt into the corrin ring, the synthesis and attachment of the dimethylbenzimidazole group via a linker arm on C17 and the attachment of an upper axial ligand labelled X, which can be a methyl- or adenosyl- group.

1.4 Enzymes of interest

1.4.1 Canonical methyltransferases

The biosynthesis of cobalamin requires eight methylation reactions to the corrin ring. The order of these methylations was discovered to be C2>C7>C20>C17>C12>C1>C15>C5 using pulse-labelling techniques[52]. Once molecular cloning became a more established technique in the nineteen nineties, this allowed the specific enzymes to be assigned including the option to overproduce the precorrin intermediates, which had been initially difficult to isolate due to their instability or low concentrations[2,52]. These are summarised in Table 1 and include the enzymes CobA, CobI, CobJ, CobM and CobF. CobL is also included but described in more depth below in section 1.4.3.

Aerobic methyltransferase (PDB code)	Addition of methyl-group to	Ancillary reaction catalysed	Anaerobic equivalent (PDB code)
CobA (not solved)	C2 & C7	None	SUMT (1S4D) CysG (1PJQ)
CobI (not solved)	C20	None	CbiL (2QBU)
CobJ (3NUT)	C17	Ring contraction	CbiH (not solved)
CobM (3NDC)	C11	None	CbiF (1CBF)
CobF (3ND1)	C1	C1 Acetic acid extrusion	CbiD (1SR8)
$CobL^{C}$ (3NjR)	C15	Decarboxylation of C12	CbiT (2YXD)
$CobL^{N}$ (not solved)	C5	None	CbiE (2BB3)
Full length CobL (not solved)	C5 & C15	Decarboxylation of C12	CbiET (to be published)

Table 1: A summary of the aerobic methyltransferases involved in cobalamin biosynthesis, the reaction(s) they carry out and anaerobic equivalents. If a structure has been solved the Protein Data Bank codes are in brackets.

All cobalamin methyltransferases are canonical (have the same fold) except $CobL^{C}$ which is a non-canonical methyltransferase and does not fit into the classification of a Class III methyltransferase like the rest. The Class III canonical methyltransferases are defined by a sequence

motif GXGXG located after the first β-strand and an active site located in a cleft between two αβα domains, in which each contains five β-strands and are parallel in the first and a mixture of parallel and antiparallel in the second, **Figure 10**[54]. In the aerobic pathway all of the methyltransferases are canonical with similar architecture, except one. CobL has both a canonical (N-terminus) and non-canonical (C-terminus) methyltransferase domain, where each adds one methyl group during the cobalamin biosynthetic pathway.

The first methyltransferase in the pathway CobA, adds two methyl-groups using the same canonical domain to catalyse both methylations to uroporphyrinogen III at C2 and C7 to form precorrin-2 (PC2)[55]. CobI is the second methyltransferase and carries out the third methylation at C20, forming precorrin-3A (PC3A)[56]. This is followed by CobG, which is not a methyltransferase and unique in the aerobic pathway because it requires oxygen to oxidise PC3A into precorrin-3B[57]. This oxidation is the set up for ring contraction and removal of C20, which is an ancillary reaction carried out after CobJ has methylated at C17, where both of these reactions result in the formation of precorrin-4 (PC4)[57]. The penultimate methyltransferase described in this section is CobM, which methylates PC4 at C11 to form precorrin-5 (PC5)[58]. Finally, CobF is the second methyltransferase of three to carry out an additional ancillary reaction, where the corrin ring decoration at C1 is removed as acetic acid, which then allows the subsequent methylation of C1 to form precorrin-6A (PC6A)[59,60].

The remaining Cob enzymes in this second phase of cobalamin biosynthesis are discussed below in more depth, including the final methyltransferase CobL. The double methylation action of CobA and CobL therefore allows the overall attachment of eight methyl-groups via only six methyltransferase enzymes. The order of these methylation reactions is important for the biochemical pathway. For example, the failure of CobI to act on C20 does not lead to successful ring contraction and the C11 methylation of CobM is vital for the isomerase action of CobH only three steps further along in the reaction pathway[56].

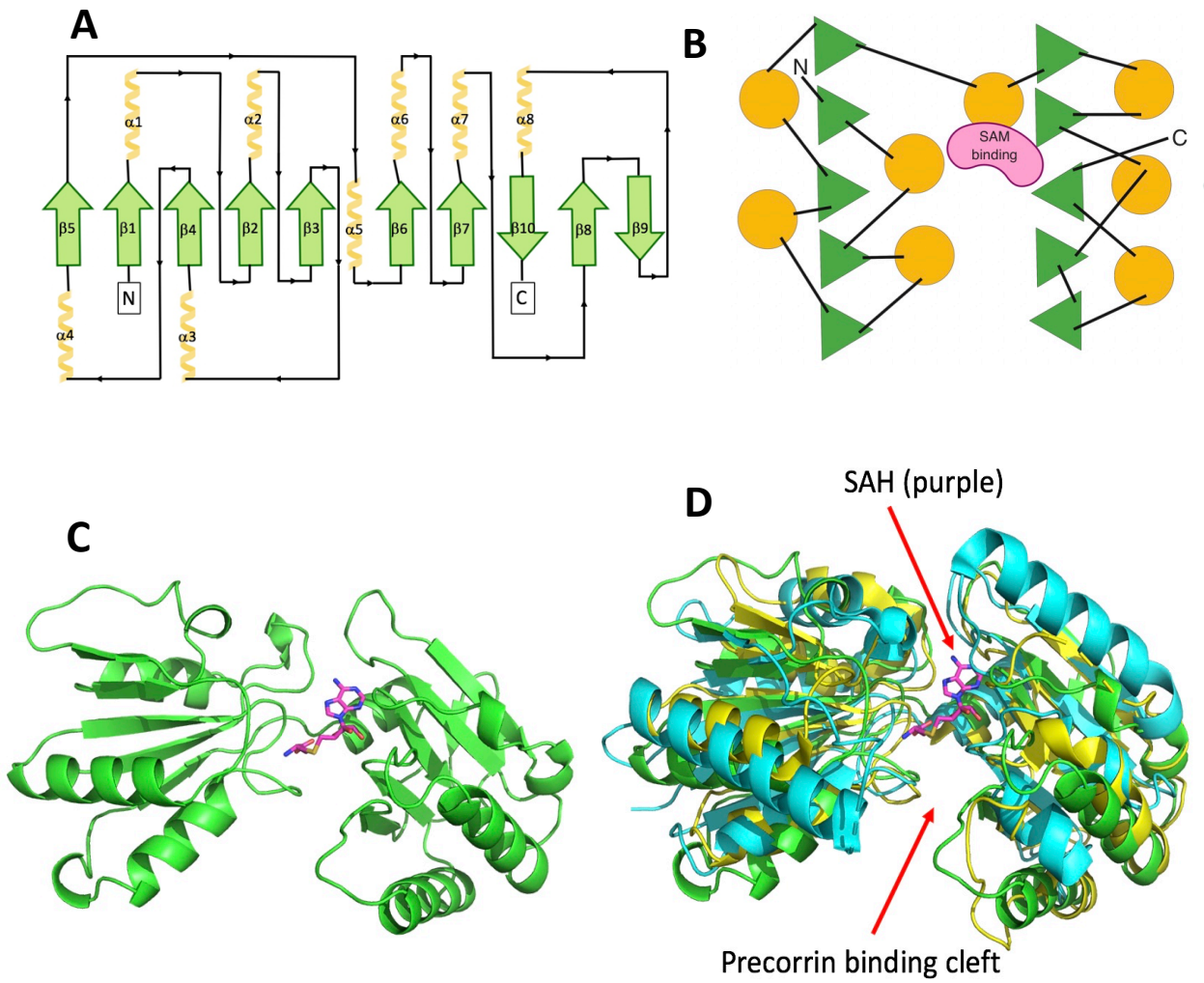


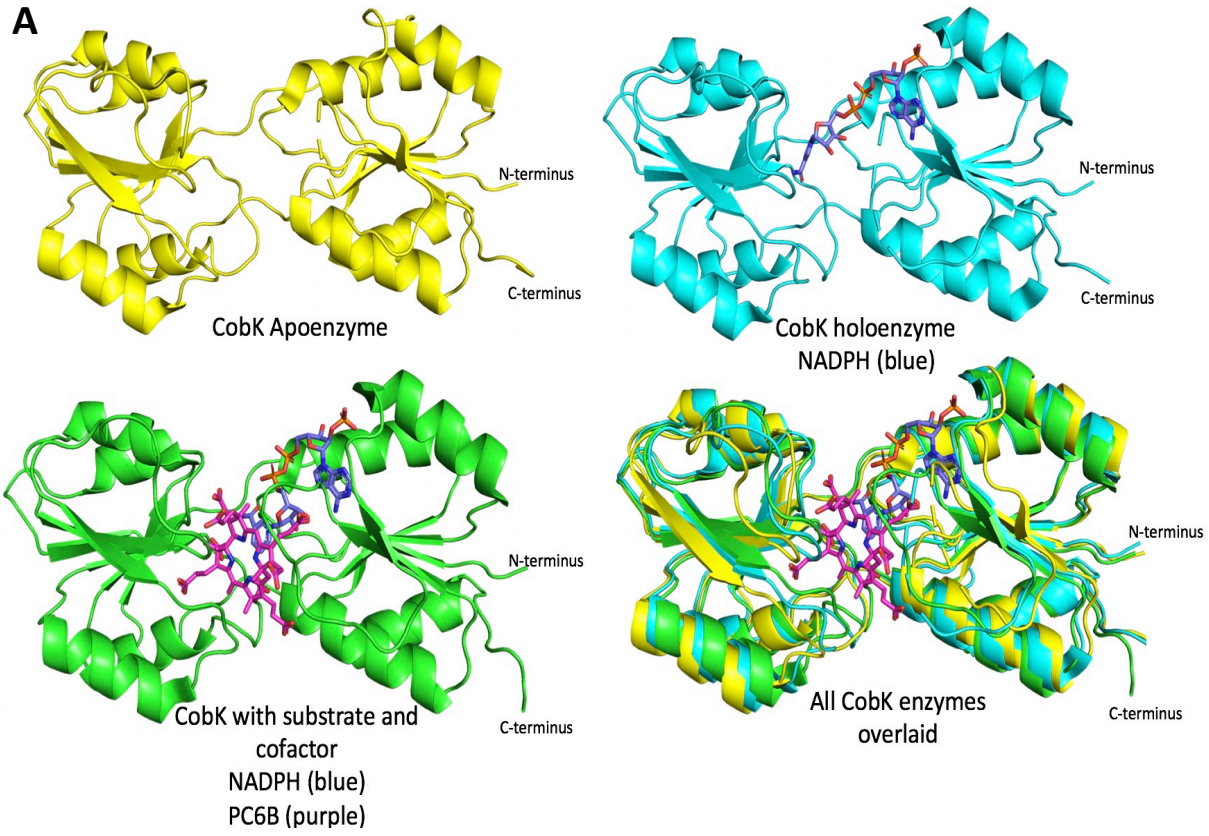
Figure 10: A) Shows a summary of the secondary structure components of Class III methyltransferases. B) Shows a simplified representation of the tertiary structure which defines Class III methyltransferases via two αβα domains in which the first domains contain five parallel β-strands. Figure adapted from Schubert et al, 2003 [54]. C) Crystal structure of CobJ monomer in green (3NUT.pdb) with S-adenosylhomocysteine (SAH) bound in purple to signify the location of the active site. D) Three methyltransferase monomers overlapped: CobJ in green 3NUT.pdb), CobM in yellow (3NDC.pdb) and CobF in blue (3ND1.pdb). Figures C and D where made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

1.4.2 CobK

Of the four Cob proteins routinely synthesised and purified in the laboratory for this project, CobK is the first in the section of the biosynthetic pathway before cobalt is inserted. Interestingly this enzyme was accidently discovered via the unexpected formation of precorrin-6B, when nicotinamide adenine dinucleotide phosphate (NADPH) was omitted in a routine conversion of uroporphyrinogen III to HBA[59]. CobK converts precorrin-6A (PC6A) into precorrin-6B (PC6B) via a reduction reaction of the carbon-18/19 double bond, in the presence of NADPH [60,61]. The numbering of this precorrin intermediate does not change as the tetrapyrrole still has six methyl groups attached and therefore two forms of precorrin-6 exist and are differentiated using A and B in the order they are respectively formed in the synthetic pathway. The crystal structure of this protein was only recently solved by the Pickersgill group and published in the Protein Data Bank (PDB)[62]. Three structures were solved; the native form and two ligand bound structures where one contained NADPH and the other contained both NADPH and the product precorrin-6B, Figure 11A. The enzyme has aroused interest as the crystal structure indicates extremely tight binding of its product via the closure of a flexible loop region over the active site. This flexible loop region was not present in the apoenzyme (5C4R.pdb) and holoenzyme structure with only NADPH bound (5C4N.pdb) but was observed closed over the active site in the structure with both NADPH and PC6B bound (5N0G.pdb). This highlights the flexibility of this post β2 loop region which is only detected clamping over the precorrin-6B ligand.

The protein structure of CobK is monomeric and can be described as two Rossmann fold domains, typically observed for nucleotide binding cofactors such as $NADP^+$ where each fold represents half of the protein[63]. Unusually these folds exchange strands between one another and are connected via a hinge like region in the middle in which the active site is located[62]. Upon substrate binding a flexible loop region closes over the active site pulling the two halves together, facilitated predominantly via conserved basic residues such as arginine and lysine. The NADPH/PC6B crystal structure shows complete encasing of the precorrin substrate/product via many binding

interactions to the ethanoic and propanoic groups of the tetrapyrrole ring, Figure 11B. The NADPH is located adjacent to the precorrin-6B but is more exposed upon binding, compared to the more complete encapsulation of precorrin-6B, as shown in Figure 11C.



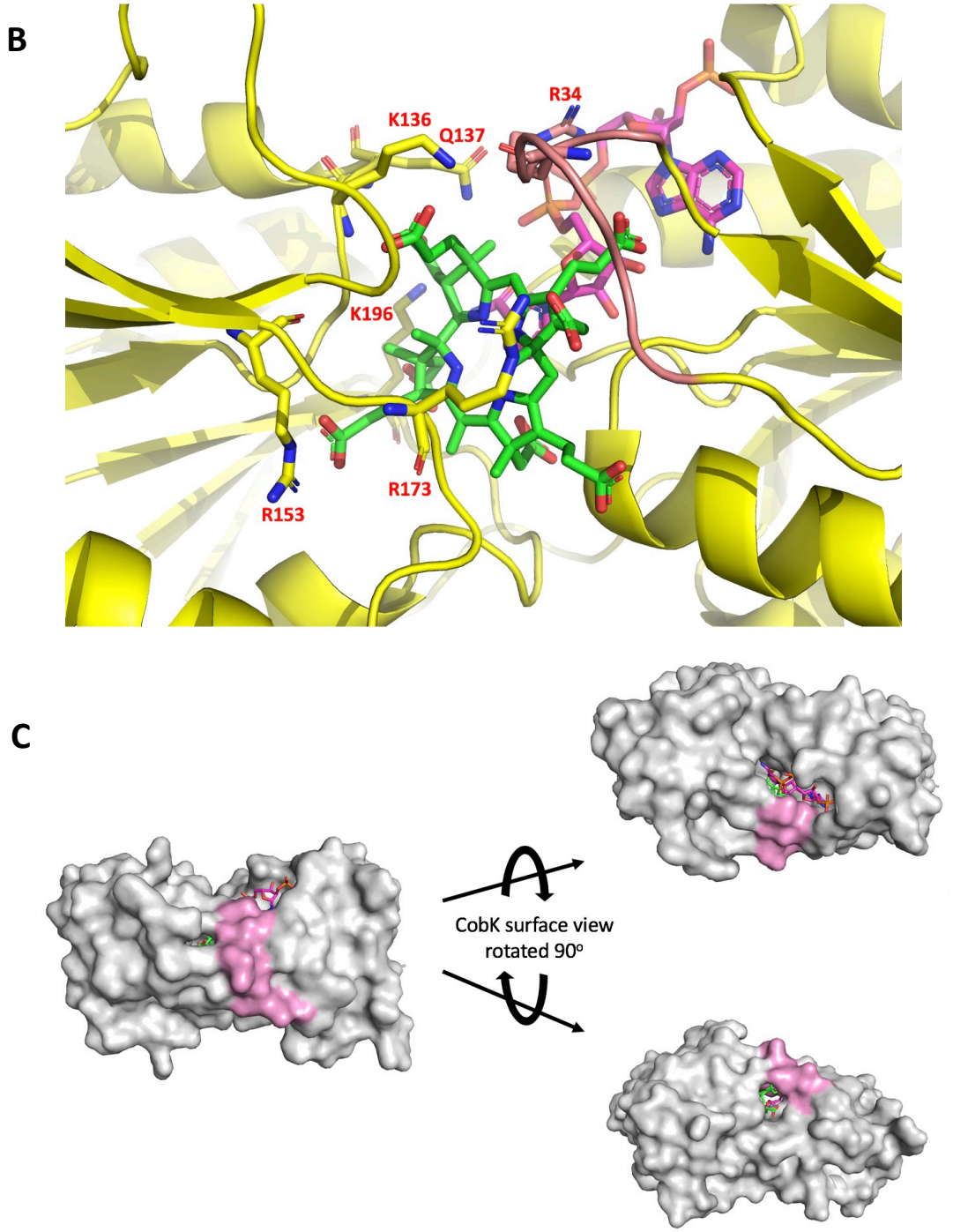


Figure 11: A) Shows a structural comparison of CobK holoenzyme (5C4R.pdb), apoenzyme (5C4N.pdb) and ligand bound form (4X7G.pbd),N.B. the PC6B was in fact methylated at C5, which is discussed in more depth later in chapter 4. B) Shows a view of the active site, where the β2/β3 loop is again highlighted in pink and key binding residues for loop closure and ligand docking are labelled with PC6B coloured in green and NADPH coloured purple. C) Shows a surface view of CobK in grey where the β2/β3 loop is highlighted in pink. This view shows how encapsulated PC6B ligand (green) is upon binding and when rotated around the active site only NADPH (purple) and a small portion of PC6B are exposed. Figures made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

1.4.3 CobL

CobL is the next enzyme in the cobalamin biosynthetic pathway after CobK. It is responsible for the conversion of PC6B into precorrin-8 (PC8), via two methylations at carbon five (C5) and carbon fifteen (C15) using s-adenosylmethionine (SAM) and a decarboxylation at carbon twelve (C12)[64]. Currently only the crystal structure of the C-terminal domain ($CobL^{c}$) has been solved (PDB-3NJR)[53]. The C-terminus is responsible for the first modification where the decarboxylation of C12 acts in tandem with the methylation, as proposed in Figure 12. This produces precorrin-7 (PC7), which is subsequently methylated at C-5 in a second active site located in the N-terminus ($CobL^{N}$).

The crystal structure of $CobL^{c}$ shows that it forms a tetramer, by way of the formation of a central β-barrel via an extended β-hairpin from each monomer[53]. The original crystal structure shows a dimer in the crystal unit cell, formed via the β-hairpin, but upon generating the crystal symmetry and protein packing in the crystal the tetramer becomes apparent, Figure 13. A remarkable structural comparison of $CobL^{c}$ with its anaerobic equivalent CbiT is demonstrated in Figure 14, highlighting the high degree of structural similarity between these non-canonical methyltransferases[65].

The anaerobic equivalent of CobL consists of the proteins CbiE and CbiT, which in some cases exists as a single protein CbiET, which is thought to have occurred via a gene fusion[65]. Individually both CbiE and CbiT have had their structures solved (PDB 1KXZ and 2BB3 respectively)[65]. CbiT shows a very strong structural similarity with $CobL^{c}$, Figure 14, while CbiE is thought to be an appropriate structural analogue to the unsolved N-terminus of CobL ($CobL^{N}$), as it carries out the same reaction of a C5 methylation. Of the methyltransferase structures solved they indicate much weaker binding affinity for tetrapyrrole binding, whereas the binding of SAM or unmethylated form S-adenosylhomocysteine (SAH) appears to be much more buried into the binding clefts. A structural comparison of the N- and C-termini of an unpublished CbiET structure is demonstrated in Figure 15. This highlights the similarity of the N-terminus to CobJ, a canonical methyltransferase in the aerobic pathway, and the C-terminus to $CobL^{C}$ a non-canonical methyltransferase of the aerobic pathway.

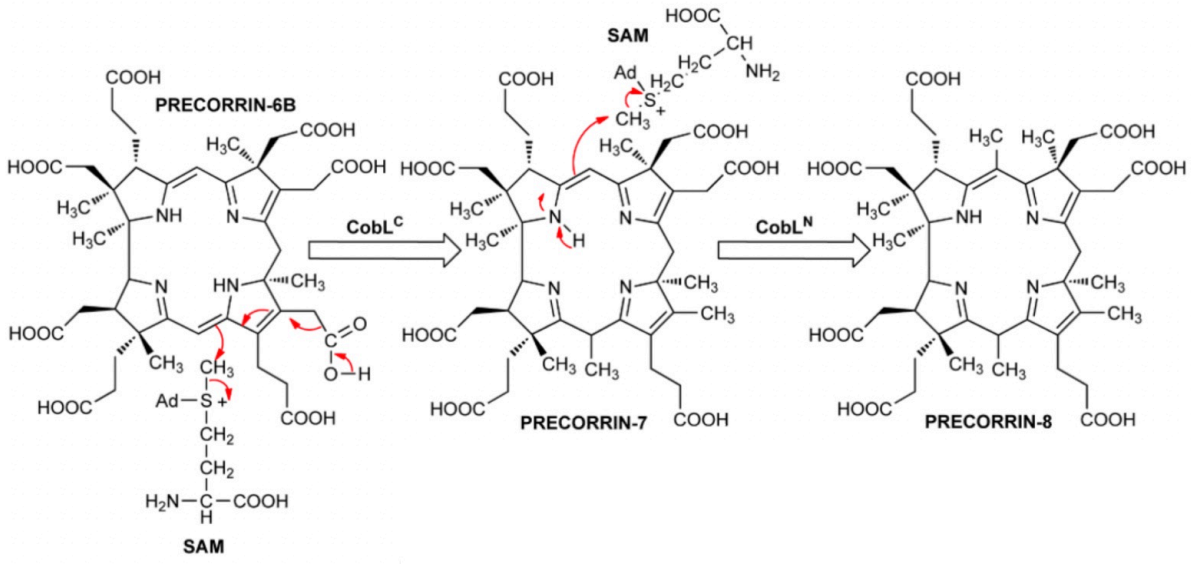


Figure 12: Proposed sequential mechanism of CobL, where the C-terminus first methylates C15 via the decarboxylation of C12. This is followed by the methylation of C5 via the N-terminal active site. Figure taken from Deery et al, 2012 [53].

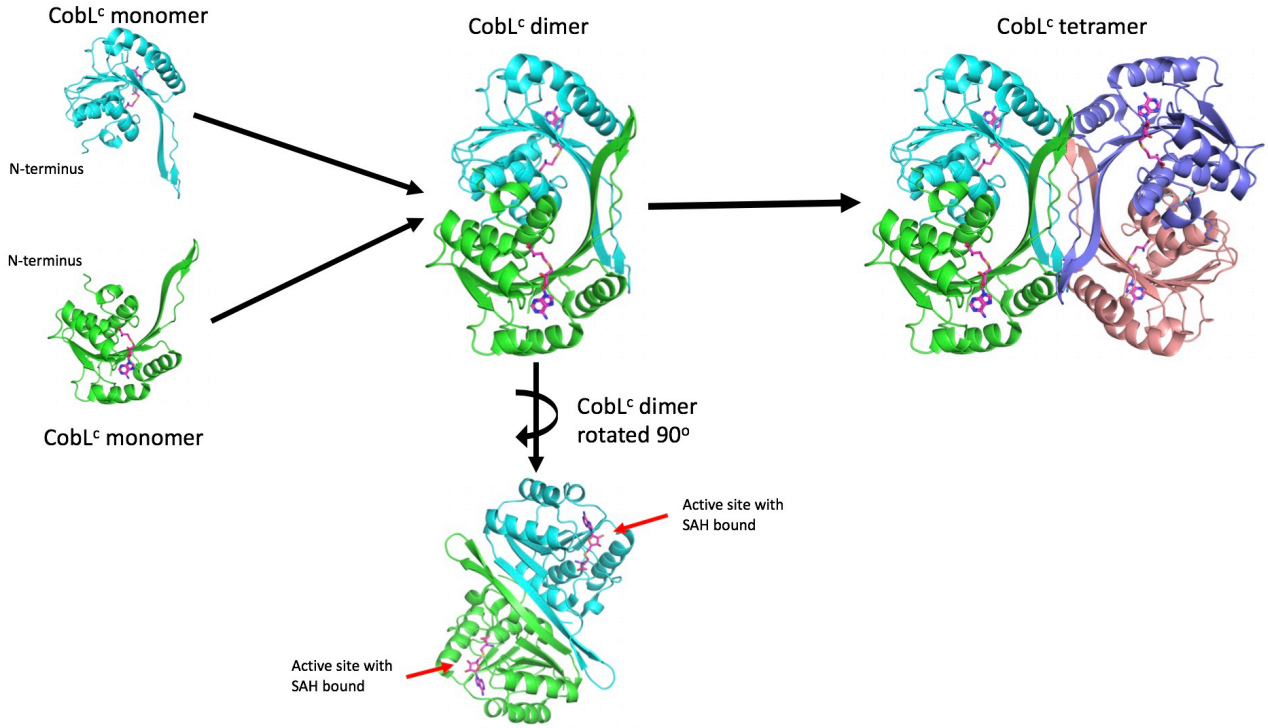


Figure 13: Shows how CobL^c monomers assemble to form a dimer or dimers. SAH is coloured in purple and signifies the location of the active site. Each monomer has a large β-hairpin present which come together to form a central β-barrel upon tetramer formation. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

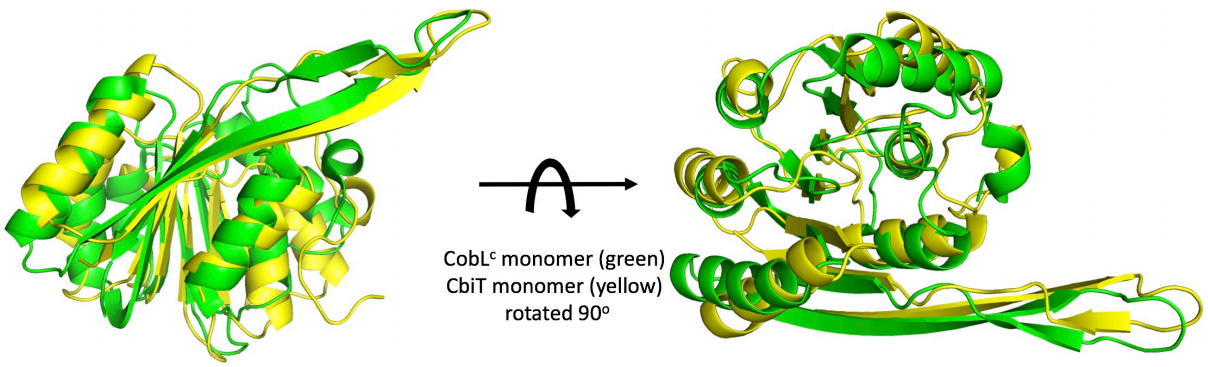


Figure 14: A structural comparison of CobL^c monomer in green (3NJR.pdb) with the anaerobic equivalent CbiT monomer in yellow (2YXD.pdb), which show a large similarity in overall tertiary structure. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

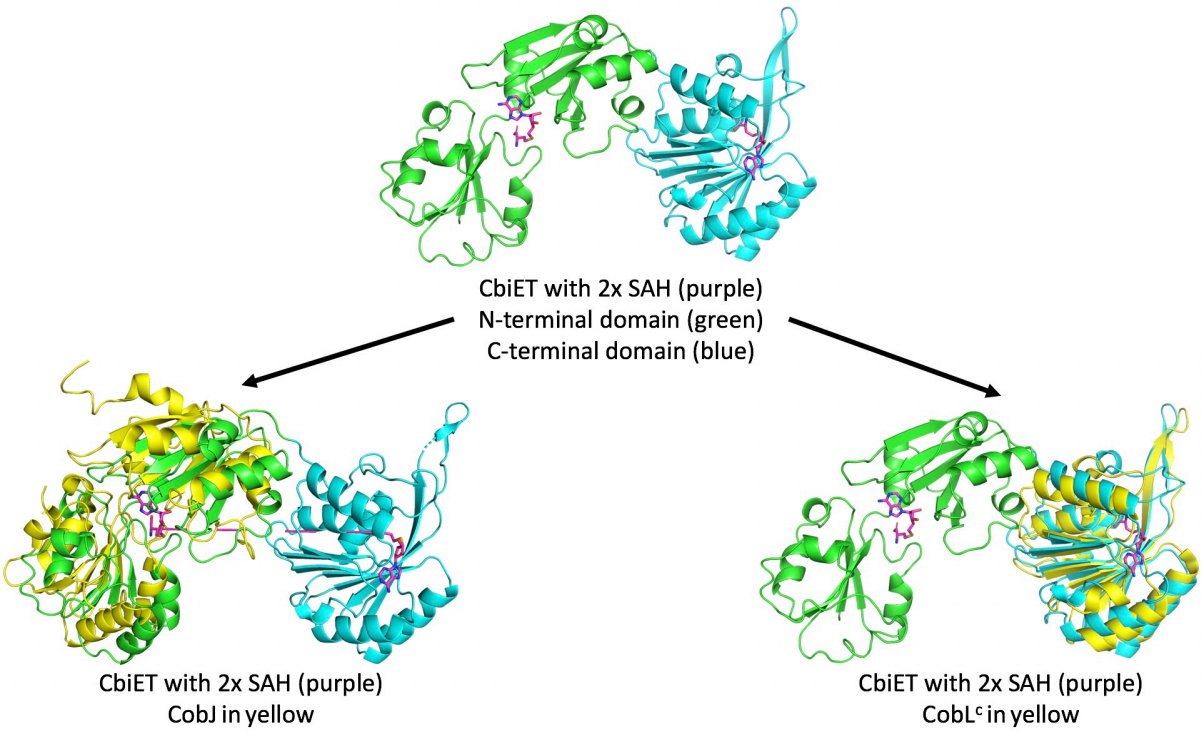


Figure 15: A structural comparison of the N- and C-termini of CbiET (unpublished) with CobJ (3NUT.pdb), a canonical methyltransferase, as described in section 1.4.1 which closely mirrors the N-terminus of CbiET (or CbiE) and also CobL^c (3NJR.pdb) which closely mirrors the C-terminus of CbiET (or CbiT). S-adenosylhomocysteine (SAH) is present in CbiET and coloured in purple, signifying the location of the two active sites present in the enzyme. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

CobL is believed to be structurally similar to CbiET, which forms a homo-tetramer, via a dimerisation of dimers[65]. Two CbiET monomers dimerise in a longitudinal fashion along the full length of the CbiE portion of the protein. This dimer then dimerises laterally via the CbiT portion of the protein where each individual monomer features two long β-strand regions which come together to form a β-barrel holding the tetramer together, where this final double dimerisation is identical to $CobL^c$ in Figure 13. There does appear to be a groove between the active sites of the C- and N-terminals, as highlighted in red in Figure 16, which may facilitate the progression of the PC7 into the second active site without loss of substrate, acting as a form of substrate channelling. This groove is purely speculative when looking at CbiET in surface view, however it is lined with charged residues such as arginine, lysine, aspartic acid and glutamic acid which would interact well with the carboxylic acid side chains of PC7 between transfer.

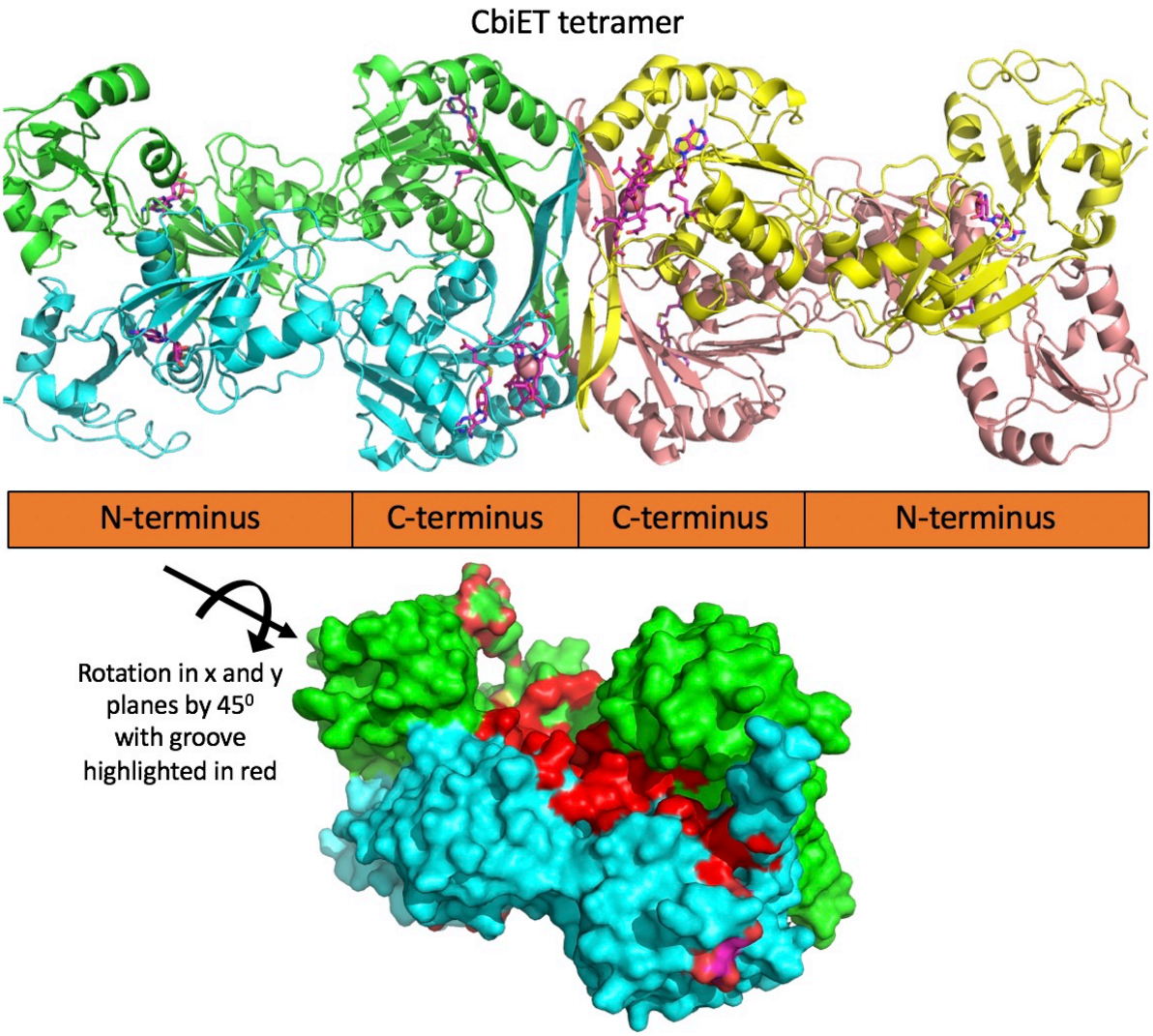


Figure 16: A structure of the unpublished CbiET tetramer. Each chain is highlighted in a different colour with SAH and cobalt-precorrin-6B both highlighted in purple, to indicate the locations of the active sites. Below is a surface view of the CbiET dimer in blue and green in which a groove highlighted in red can be observed between the active sites of the N- and C-termini of separate protein chains. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

1.4.4 CobH

CobH is the third enzyme of interest in the biosynthetic pathway, following on from CobL. In a similar fashion to CobK it has aroused interest due to unusual product binding. CobH catalyses the isomeric rearrangement of precorrin-8 into hydrogenobyrinic acid (HBA), by moving the methyl-group located at C11 to C12 via a sigmatropic rearrangement reaction[66,67]. Figure 17A shows a proposed mechanism for CobH via a catalytic histidine residue. This reaction causes a highly visible colour change from yellow to pink/red of the cobalamin intermediate.

CobH is a homodimer where the two active sites are formed between the two monomers. The crystal structures have been solved (PDB 1I1H and 1F2V)[67], where again the structure supports product binding. Tight binding is a result of the complete burial of ligand and numerous binding interactions, via basic residues, such as charged arginine or uncharged polar groups such as serine, asparagine and glutamine, Figure 17B.

To support the observed product binding activity Thibaut et al, 1992 also observed that CobH had an extremely slow catalytic turnover with a K_{cat} of $5.1h^{-1}$ which corresponded to a V_{max} of $230nmolh^{-1}mg^{-1}$. They took this one step further and showed that when CobH is in the presence of HBA it is inhibited through competitive means. This is not surprising given CobH is a mutase and simply rearranges the substrate into a slightly altered isomeric form. An interesting side study from this paper found that HBA was able to inhibit the production of the anaerobic equivalent cobyrinic acid, whereas hydrogenobyrinic acid-a,c-diamide produced by the subsequent enzyme CobB does not, highlighting that the a- and c- chains must play a role in dictating specificity in CobH binding. The protein face around the active sites appears to have a rather flat edge, when viewed in surface mode, which may provide a protein-protein interaction site for assisting in correctly delivering PC8 or removal of HBA.

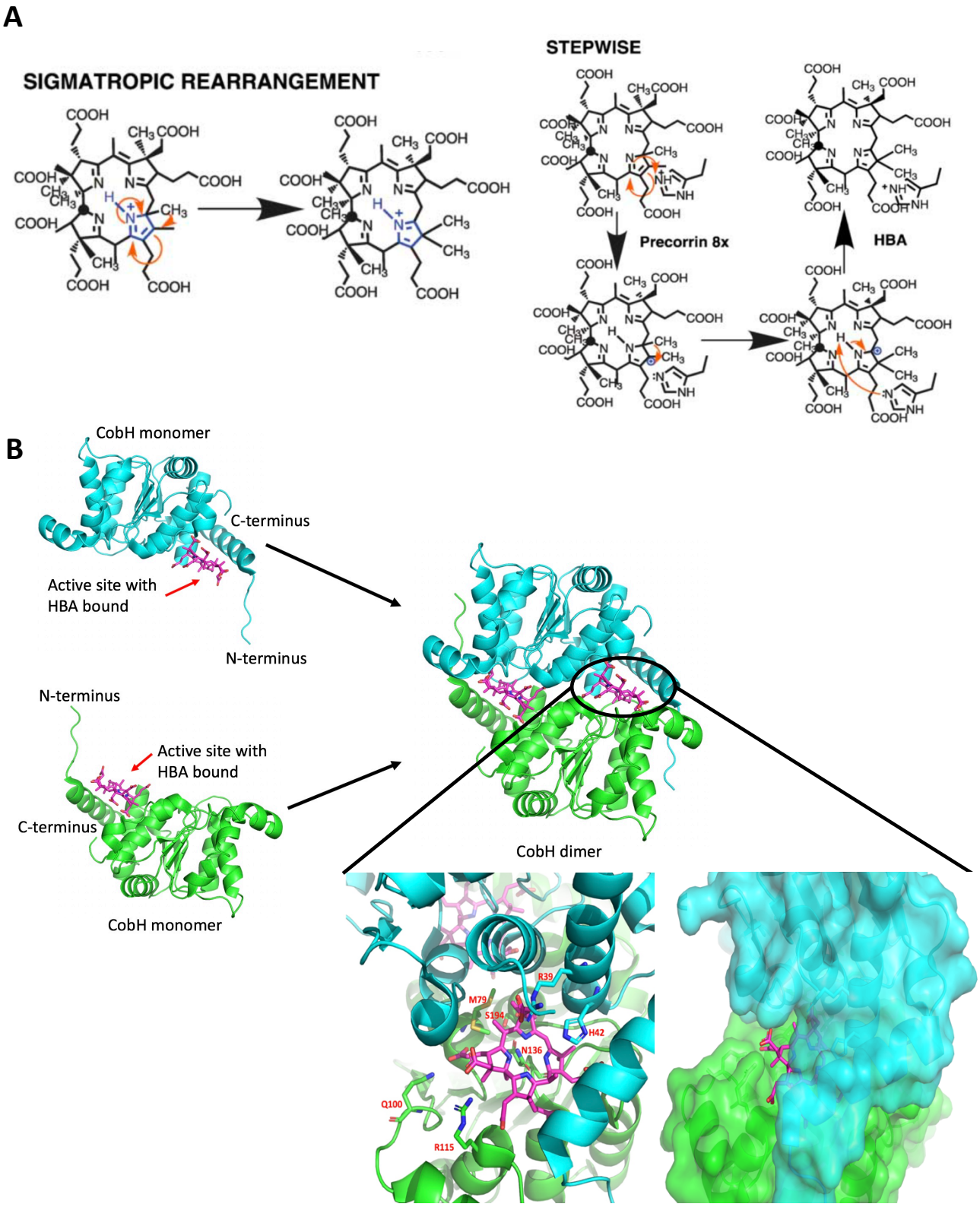


Figure 17: A) Proposed mechanism taken from Shipman et al,2001 [67] showing the simatropic rearrangement involved for the formation of HBA via a catalytic histidine residue. B) Shows how CobH (4FDV.pdb) dimer formation comes about, with HBA coloured in purple, and includes a close up of the active site, to highlight interacting binding residues and a surface view to show almost complete burial of HBA. Figure B made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

1.4.5 CobB

CobB is the final enzyme of interest in this study and catalyses the reaction after CobH, where hydrogenobyrinic acid is amidated to hydrogenobyrinic acid-a,c-diamide (HBAD). The di-amidation takes place on two of the seven carboxyl groups, designated *a* and *c* on the porphyrin ring [51]. CobB currently has not had its crystal structure solved. Galperin and Grishin[68] carried out an in-depth sequence analysis of CobB and concluded that it belongs to a family of enzymes known as class I glutamine-dependent amidotransferases. This class of enzymes typically have a low sequence identity however they do exhibit a strong structural conservation[68,69,70]. These enzymes contain a glutamine binding domain and an ATP binding domain. Class I glutamine-dependent amidotransferases are identified via their catalytic triad consisting of cysteine, histidine and glutamic acid, which are required for glutamine binding. The glutamic acid and histidine residues assist in the activation of the cysteine thiol, which cleaves the amine group from the glutamine substrate[71]. Surprisingly CobB lacks the glutamic acid residue of the catalytic triad, however this is thought to be the least important of the three catalytic residues[70]. ATP binding occurs in the ATP binding domain which can be typically defined by the well-defined P-loop fold[72]. The P-loop contains two conserved sequence motifs, the Walker A motif which is responsible for positioning the triphosphate, which typically has the pattern GXXXXGK, and the Walker B motif which is normally an aspartate (or glutamate) residue involved in the binding of a water-bridged magnesium ion[73,74].

The mechanism of action in class I amidotransferases occurs through a coupled mechanism between the separate active sites. The glutaminase activity only occurs when all substrates are bound, thus avoiding any wasteful hydrolysis and loss of substrates[73,75]. In CobB the substrate HBA is phosphorylated via ATP twice where it has been shown that the carboxyl chain assigned *c* is amidated before the carboxyl chain assigned *a* on the corrin ring[51]. The amine group is donated via glutamine in the form of ammonia, which then travels through an intramolecular tunnel. This avoids protonation into ammonium, where it can then amidate the substrate via nucleophilic attack[76]. Figure 18 shows the mechanism of how class I glutamine-dependent amidotransferases bind

glutamine and how a basic residue, typically histidine activates the thiol group of the catalytic cysteine. Thereafter the cleaved ammonia is transferred to a separate site upon which it amidates a separate substrate via an S_N1 reaction[76,77].

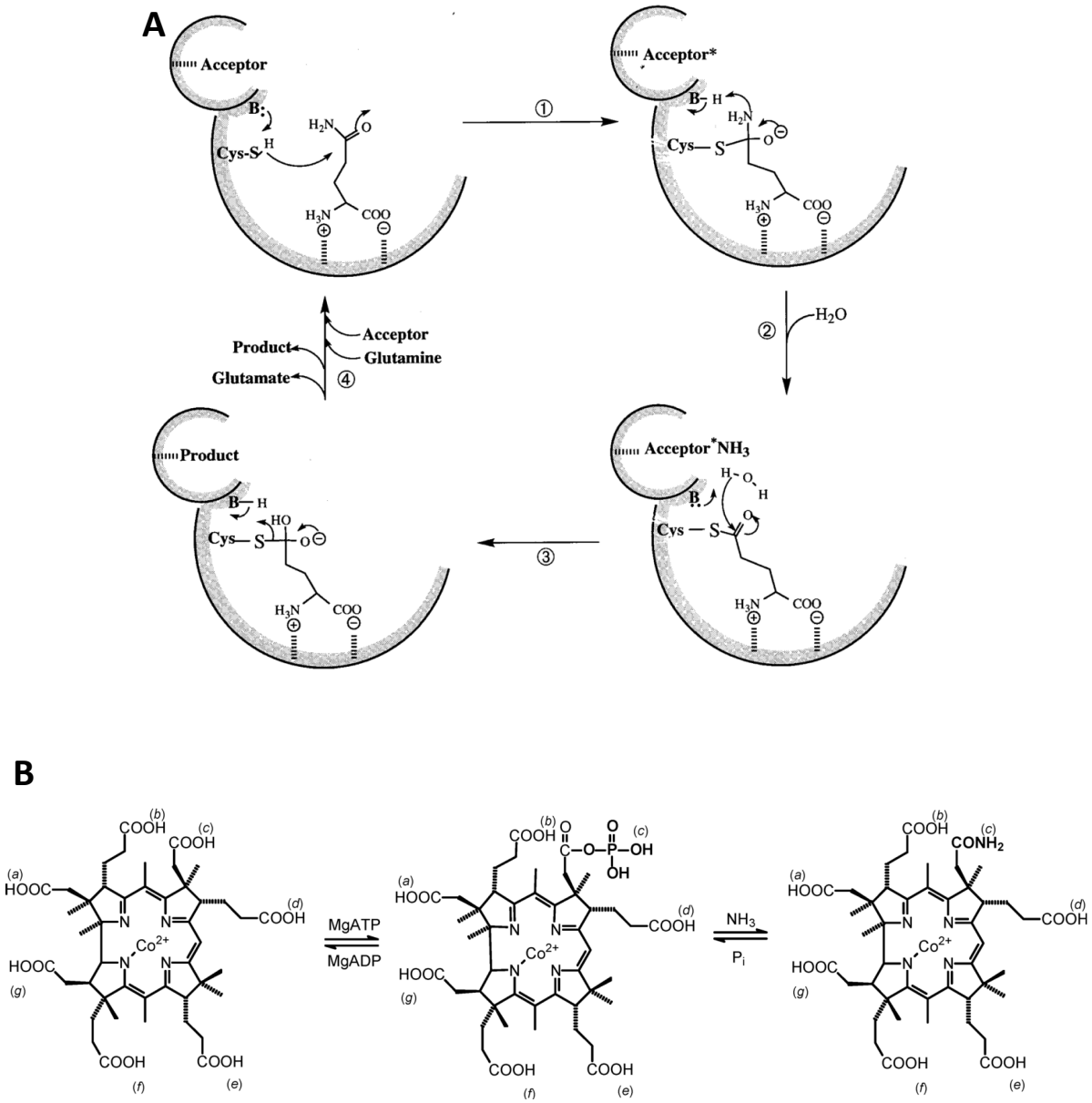


Figure 18: A) Shows the mechanism of action of the glutamine binding domain of CobB, where it can be seen that an activated cysteine residue carries out a nucleophilic attack to cleave the amine group of glutamine. Figure taken from Massiere et al, 1995 [70]. B) Shows the reaction sequence of the ATP binding domain of CobB, where ATP first phosphorylates the C7 c-chain of HBA followed by a nucleophilic substitution by ammonia. Figure taken from Fresquet et al, 2004 [78].

Many amidotransferases have been shown to contain molecular tunnels, typically ranging between 20-30Å, which serve to protect intermediates, enhance catalysis or reduce transient time[77,79–81]. Figure 19 shows varying representations of CobB. Figure 19B Is the crystal structure of CTP synthase, which carries out the regulation of pyrimidine levels by converting UTP to CTP via an amidation reaction[82]. CTP synthase is a good comparison as it is a full-length protein containing both a class I glutamine amidotransferase domain and an ATP binding domain required for CTP production. Other protein models include glutamate synthase, GMP synthetase, MinD and dethiobiotin synthetase to name a few, but these only contain one of the domains found in CobB[70,71,73,83,84]. Finally, Figure 19C shows a predicted model of the ATP binding domain of CobB by Galperin and Grishin, 2000[68].

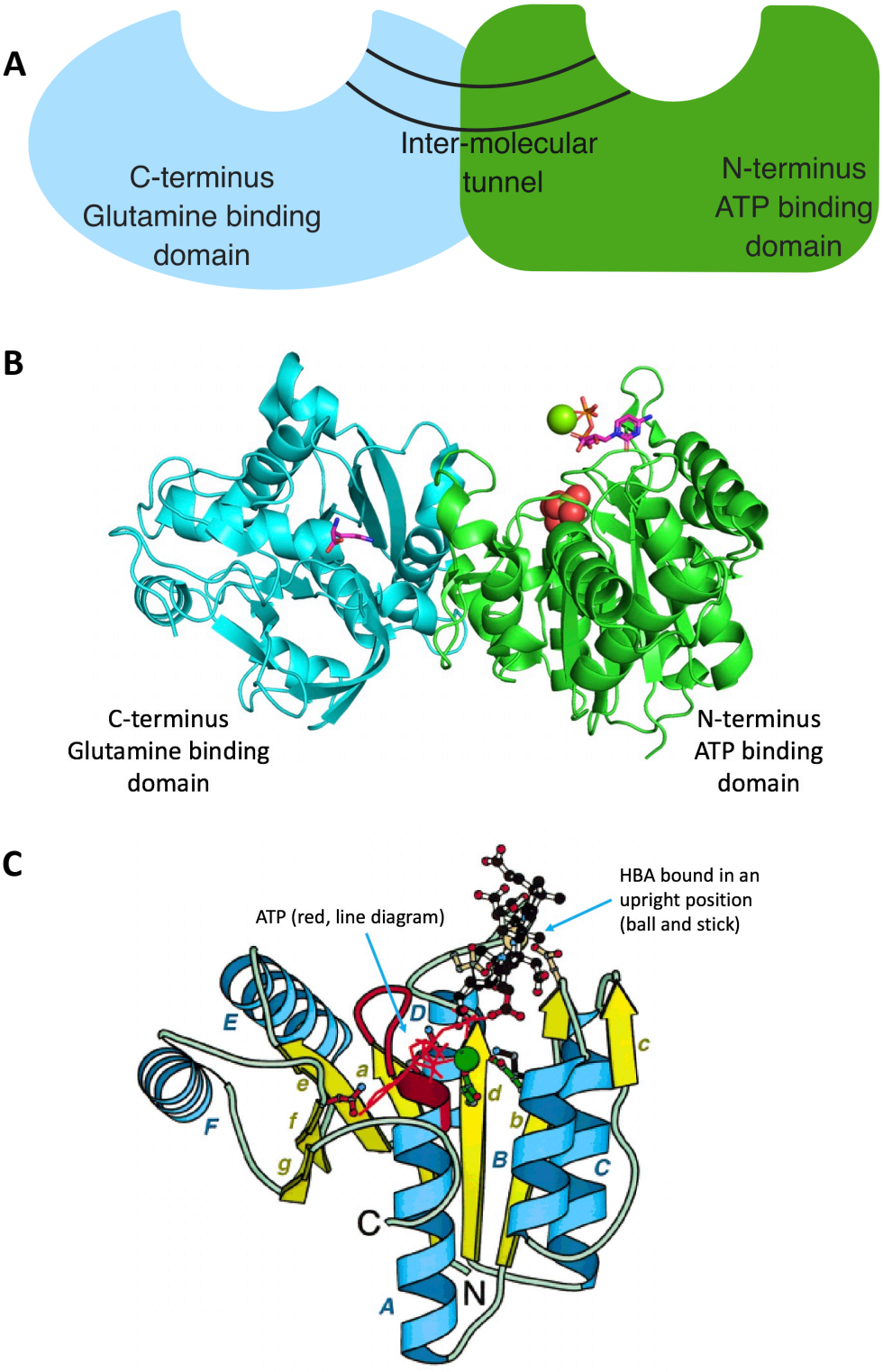


Figure 19: A) Shows a simplified model of CobB showing the glutamine and ATP binding domains, created using Biorender.com. B) Shows CTP synthase (5TKV.pdb) with the substrates glutamine and CTP coloured in purple. CTP is associated with a Mg ion (green sphere). The ATP domain also contains two sulphate ions which is most likely where ATP binds. C) Shows a computational model of the ATP binding domain taken from Galperin and Grishin, 2000[68]. Figure B made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

The anaerobic equivalent of CobB is CbiA, which has had an in-depth study on it carried out by Fresquet et al, 2004 [78] who summarised that CbiA acts as a monomer, whereas CobB is a dimer. Unsurprisingly their study of CbiA found it has a 500-fold higher affinity for its substrate cobyrinic acid, than the aerobic equivalent substrate HBA from the aerobic pathway. They also demonstated that ammonia can be substituted as a nitrogen source which was also observed originally by Debussche et al 1990[51]. However, Fresquet et al, 2004[78] demonstated that glutamine hydrolysis can occur in the absence of other substrates but is enhanced in the presence of ATP or cobyrinic acid. This is assumed to be coordinated by a conformational change that is transmitted allosterically upon substrate binding. Both active sites can act independently either phosphorylating cobyrinic acid or glutamine cleavage but CbiA works best when all active sites are occupied. Surprisingly even at saturating concentrations it was found that these two active sites were not completely coupled to each other, with the rate of glutamine hydrolysis being around eight times faster than ADP production. However, Fresquet et al, 2004 were unable to explain how the sequential order of the amidations is co-ordinated, given the cobyrinic acid needs to leave the active site and re-binds after rotating.

1.5 Purpose of thesis

The biosynthesis of cobalamin has been studied for decades. The recent crystal structure of CobK[62] solved in the Pickersgill group and the enzyme trap study[53] have prompted the idea of product binding activity in both of the enzymes CobK and CobH. This activity raises questions such as: (i) Does this activity help or hinder the pathway? (ii) How are the products released? (iii) How is selectivity for such similar substrates brought about?

With a particular focus on looking at how the subsequent enzyme in the pathway receives its substrate, an understanding of this unusual product binding activity may help to further clarify some of the complexity and coordination involved in the biosynthesis of cobalamin. Within this thesis a

number of approaches are described that have been used to attempt to probe product trapping activity of cobalamin biosynthetic enzymes further. In Chapter 3, the individual protein purifications for X-ray crystallography are described for CobL, CobB and co-crystallography attempts of the enzyme pairs CobK/L and CobH/B. In Chapter 4, product binding activity of PC6B to CobK is demonstrated using UV spectroscopy, isothermal calorimetry (ITC) and nuclear magnetic resonance (NMR) and concludes with techniques including X-ray crystallography and fluorescence spectroscopy which show incorrect cobalamin intermediates binding to CobK. This prompted an in-depth analysis of the canonical methyltransferase structures in Chapter 5, where their structures are described and ligand docking is explored to try and elucidate how specific precorrin binding is brought about. To probe these research questions further the main aims of the project were:

1. To solve the structures of full length CobL and CobB using crystallography to provide further insights on how these enzymes obtain their substrates from the product binding enzymes CobK and CobH.

2. To further explore the product binding activity and release of PC6B to CobK so as to determine why CobK may do this and how CobL, the subsequent enzyme in the pathway, obtains its substrate.

3. To provide a possible explanation into how the stereo-specificity of the cobalamin biosynthetic enzymes are able to differentiate between different cobalamin intermediates, with a particular focus on the canonical methyltransferases, due to a large amount of structural data available on these enzymes.

Chapter 2: Materials and Methods

2.1 Chemicals

Unless otherwise stated all chemicals and reagents were purchased from Sigma-Aldrich, which is now Merck. Cobalamin intermediates were made in the laboratory as described in section 2.3.4.

2.2 Genes

The five main proteins of interest produced in the laboratory for this project were CobK, CobL (full length), $CobL^c$ (the C-terminal domain of CobL), CobH and CobB. These enzymes are present in the aerobic production of cobalamin and listed in the sequential order required for cobalamin biosynthesis. The gene sequences used were all sourced from *Rhodobacter capsulatus*.

For the methyltransferase sequence review in Chapter 5, sequences came from proteins that already have known structures and are available in the Protein Data Bank. A summary of all the protein sequences used in this project are listed in Table 2, alongside the protein data bank codes used for structural studies.

Protein	Organism	PDB Code
CobA	*Rhodobacter capsulatus*	-
CysG	*Salmonella typhimurium*	1PJQ
SUMT	*Sinorhizobium*	1S4D
CobI	*Rhodobacter capsulatus*	-
CbiL	*Methanothermobacter thermautotrophicus*	2QBU
CobJ	*Rhodobacter capsulatus*	3NUT
CbiH	*Bacillis megaterium*	-
CobM	*Rhodobacter capsulatus*	3NDC
CbiF	*Bacillis megaterium*	1CBF
CobF	*Rhodobacter capsulatus*	3ND1
CbiD	*Archaeoglobus fulgidus*	1SR8
CobK	*Rhodobacter capsulatus*	4X7G
CobL (C-terminus)	*Rhodobacter capsulatus*	3NJR
CbiE	*Archaeoglobus fulgidus*	2BB3
CbiT	*Methanothermobacter thermautotrophicus*	1L3I
CobH	*Rhodobacter capsulatus*	4FDV
CobB	*Rhodobacter capsulatus*	-

Table 2: A summary of all the cobalamin biosynthetic proteins used in this project, along with the Protein Data Bank code if they have had their structure solved.

2.3 Microbiology techniques

2.3.1 Media for microbial growth

Different media and antibiotics were used throughout the project and their compositions are described in Table 3.

Chemical	Chemical components
LB media	10g/L Trypton 5g/L Yeast Extract 5g/L NaCl Autoclaved
LB Agar	LB media 15g/L Agar Autoclaved
2TY media	16g/L Trypton 10g/L Yeast Extract 5g/L Sodium Chloride Autoclaved
Minimal Media (1L)	100ml of M9 Salts 10ml of 100x stock trace elements 840ml of autoclaved dH_2O 50ml of filtered sterile additives Add appropriate antibiotics
10x M9 Salts	30g/L KH_2PO_4 120g/L Na_2HPO_4 5g/L NaCl Autoclaved

Chemical	Chemical components
100x Stock trace elements (200ml)	1g EDTA 166mg $FeCl_3.6H_2O$ 3mg $ZnCl_2$ 3mg $CuCl_2.2H_2O$ 2mg H_3BO_3 2mg $CoCl_2.6H_2O$ 0.32mg $MnCl_2.6H_2O$ Adjusted pH to 7.5 with NaOH Added dH_2O to 200ml
Chloramphenicol stock	34mg/ml in ethanol Filter sterilised Stored at -20^{0}C
Ampicillin stock	100mg/ml in dH_2O Filter sterilised Stored at -20^{0}C
Filtered Sterile Additives (200ml)	4ml of 1mg/ml thiamine 4ml of 1mg/ml biotin 1.2ml of autoclaved 1M $CaCl_2$ 4ml of autoclaved 1M $MgSO_4.7H_2O$ 80ml of 20% w/v glucose 4g of NH_4Cl Added dH_2O to 200ml Filter sterilise through 0.02µm filter
-	-

Table 3: Summary media and related chemical components required for protein production in this project.

2.3.2 Cloning and isolation of plasmid DNA

At the start of this project all individual proteins were already present within pET14b plasmids, which encoded an N-terminal histidine tag; they were ready to transform for laboratory use. These constructs had been prepared by Dr Evelyne Deery who is currently based at the University of Kent[53].

Fresh plasmid stocks of CobK, CobL, CobL^{c}, CobH and CobB were prepared by transforming into DH5α™ Competent Cells. This was undertaken by adding 5-10ng of plasmid DNA to 50µl of DH5α™ Competent Cells which were then left on ice for thirty minutes. After this time the cells were then heat shocked in a water bath set at 42^{0}C for forty-five seconds before being placed back on ice for two minutes. Following this 950µl of LB media was then added to the cells which were then incubated at 37^{0}C and shaken at 200 rotations per minute (rpm) for one hour. The 1ml of DH5α™ competent cell culture was then centrifuged for two minutes at 12,000g. This resulted in the cells being pelleted and 700µl of supernatant was removed before resuspending the pellet. To allow for a tenfold dilution 200µl and 20µl of cells were spread out on to separate LB agar plates which contained 100µg/ml of the antibiotic ampicillin. These plates were left overnight in an incubator at 37^{0}C. After twenty-four hours a single cell colony was picked and inoculated to a 5ml LB culture which contained 100µg/ml of ampicillin and this was left to grow at 37^{0}C overnight. After twelve hours a QIAprep® Miniprep was carried out using the solutions and instructions provided:

1. Resuspend pelleted bacterial cells in 250µl Buffer P1 and transfer to a microcentrifuge tube.
2. Add 250µl Buffer P2 and mix thoroughly by inverting the tube 4-6 times.
3. Add 350µl Buffer N3. Mix immediately and thoroughly by inverting the tube 4-6 times.
4. Centrifuge for 10 mins at 13,000rpm in a table top microcentrifuge.
5. Apply 800µl of the supernatant from step 4 to the QIAprep 2.0 spin column by pipetting.
6. Centrifuge for 30-60s. Discard the flow-through.

7. (Recommended) Wash the QIAprep 2.0 spin column by adding 0.5ml Buffer PB and centrifuging for 30-60s. Discard the flow-through.
8. Wash QIAprep 2.0 spin column by adding 0.75ml Buffer PE and centrifuge for 30-60s.
9. Discard the flow-through and centrifuge at full speed for an additional 1 min to remove residual wash buffer.
10. Place the QIAprep 2.0 column in a clean 1.5ml microcentrifuge tube. To elute DNA, add 50µl Buffer EB or water to each QIAprep 2.0 spin column, leave to stand for 1 min and centrifuge for 1 min.
11. The pure plasmid DNA was stored at -20^{0}C until required.

To prepare plasmid stocks of tetrapyrrole producing cells the same method as above was followed but it typically required a longer incubation on agar plates to produce cell colonies, due to the large size of the plasmids, Figure 20. These constructs were again made by Dr Evelyne Deery via the link and lock method[53]. These plasmids consist of all the aerobic Cob proteins up to the Cob protein able to synthesise the desired tetrapyrrole product, in which the final enzyme in the sequence contained a histidine tag. These constructs include; a precorrin-6B producing plasmid (all Cob proteins up to CobK), a hydrogenobyrinic acid producing plasmid (all Cob proteins up to CobH) and hydrogenobyrinic acid-a,c-diamide producing plasmid (all Cob proteins up to CobB).

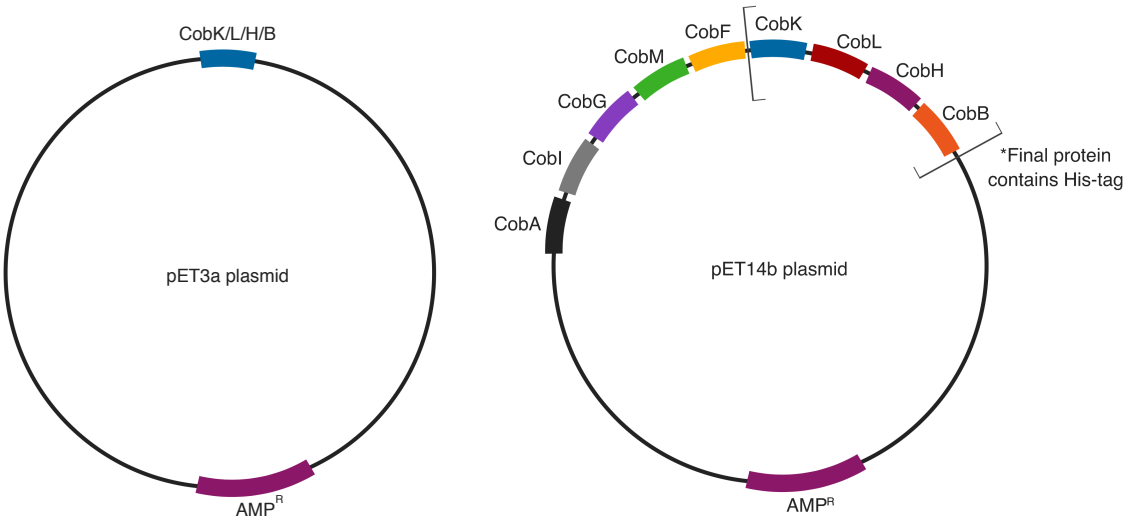


Figure 20: The different plasmids used for Cob protein and cobalamin intermediate production. A) Shows a pET3a plasmid used for the production of individual Cob proteins. These plasmids contained a T7 RNA promoter, TEV cleavage site and an N-terminal His-tag on the protein of interest as well as an ampicillin resistance gene. B) Showed the pET14b plasmid produced by Dr Evelyne Deery containing all the Cob biosynthetic proteins up to the desired cobalamin intermediate shown in the brackets, where the final protein also contained an N-terminal histidine tag. Figure was created using Biorender.com.

2.3.3 Production of recombinant proteins

For production of an individual Cob protein, plasmid DNA was transformed into One Shot™ BL21 (DE3) Chemically Competent *E. coli* cells. The transformation procedure was identical to section 2.3.2 above, but after cells were grown overnight on agar plates, colonies were picked and inoculated into 10ml LB cultures containing 100µg/ml ampicillin and grown at 37^0C shaken overnight at 200rpm. After twelve to sixteen hours the 10ml cultures were inoculated into one litre LB media batches, which again contained 100µg/ml of ampicillin. The one litre cultures were then grown at 37^0C and shaken at 220 rpm until an OD_{600} reading of 0.6-0.8AU was obtained, which typically took around three hours. At this point 0.4mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. Cell cultures were then grown for a further eighteen hours at 18^0C before the cells were harvested by centrifugation at 8,000g for fifteen minutes. The supernatant was poured off and the cell pellets were collected and stored at -20^0C.

2.3.4 Growth of cobalamin intermediate producing cells

The cobalamin intermediate producing plasmids: precorrin-6B (PC6B), hydrogenobyrinic acid (HBA) and hydrogenobyrinic acid-a,c-diamide (HBAD) were transformed into One Shot™ BL21 (DE3) pLysS chemically competent cells as described in section 2.3.2 above. However, the agar plates required the addition of 34µg/ml chloramphenicol in addition to 100µg/ml ampicillin due to the pLysS host cells to reduce basal protein expression. Cell colonies were picked and inoculated into 10ml 2TY cultures, which contained 34µg/ml and 100µg/ml chloramphenicol and ampicillin respectively and grown overnight at 37^0C and shaken at 200rpm. After twelve to sixteen hours the 10ml cultures were transferred into one litre 2TY media batches, which again contained 34µg/ml chloramphenicol and 100µg/ml ampicillin and grown at 37^0C until an OD_{600} reading of 0.6-0.8AU was obtained, which typically took around three hours. At this point 0.4mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. After induction the temperature was lowered to 18^0C and the batches were shaken at 220rpm for 24-36hrs. Typically this would produce a greyish-green coloured culture for PC6B producing cells and pink coloured cultures for HBA and HBAD producing cells. Cells were harvested by centrifugation at 8,000g for fifteen minutes. The supernatant was removed, and the cell pellets collected and stored at -20^0C.

2.4 Purification of recombinant proteins

2.4.1 Protein purification buffers

All individual Cob proteins contained a histidine tag, so a nickel column followed by a size exclusion chromatography using Superdex 200 Increase column a was used in tandem for protein purification. Depending on the protein being purified the purification buffers used varied and are summarised in Table 4.

Cob Protein	Loading buffer (Buffer A)	Wash buffer (Buffer B)	Elution buffer (Buffer C)	Desalting buffer (Buffer D)
CobK	100mM NaCl 20mM HEPES pH7.5 5mM imidazole	100mM NaCl 20mM HEPES pH7.5 40mM imidazole	100mM NaCl 20mM HEPES pH7.5 400mM imidazole	100mM NaCl 20mM HEPES pH7.5
CobL	500mM NaCl 20mM Tris-HCl pH8.0 5mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0 60mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0 400mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0
CobL^{c}	500mM NaCl 20mM Tris-HCl pH8.0 5mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0 60mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0 400mM imidazole	500mM NaCl 20mM Tris-HCl pH8.0
CobH	100mM NaCl 20mM Hepes pH7.5 5mM imidazole	100mM NaCl 20mM Hepes pH7.5 60mM imidazole	100mM NaCl 20mM Hepes pH7.5 400mM imidazole	100mM NaCl 20mM Hepes pH7.5
CobB	100mM NaCl 20mM Hepes pH7.5 5mM imidazole 2mM TCEP Kept on ice	100mM NaCl 20mM Hepes pH7.5 60mM imidazole 2mM TCEP Kept on ice	100mM NaCl 20mM Hepes pH7.5 400mM imidazole 2mM TCEP Kept on ice	100mM NaCl 20mM Hepes pH7.5 2mM TCEP Kept on ice

Table 4: Summary of protein purification buffers used for each Cob protein.

2.4.2 Lysis of cells

Typically, 1-2 cell pellets (~5-10g) were used for each protein purification, where each pellet had originally come from a one litre cell culture. The cell pellets were left on ice to thaw for thirty minutes before being resuspended in 30ml of buffer A. The resuspended cells were then lysed by sonication on ice for six minutes with thirty second intervals at 25-30 watts. The cell lysate was then centrifuged at 25,000g for thirty minutes at 4^0C and the supernatant was collected and stored on ice.

2.4.3 Purification of His-tagged proteins

The supernatant from the previous steps was then loaded on to a nickel column, which had been washed with five column volumes of loading buffer (buffer A). Once all the supernatant had been passed through the column, five column volumes of buffer A was washed through the nickel column. This was then followed by a second wash of five column volumes of wash buffer (buffer B). Finally, two column volumes of elution buffer (buffer C) were added to the nickel column and collected, as this elution contained the protein of interest.

2.4.4 Protein buffer exchange

The nickel column elution was concentrated down to 2.5ml in a Vivaspin® 20 centrifugal concentrator with an appropriate molecular weight cut off (MWCO), depending on the protein being purified. This was typically a 10kDa MWCO or 30kDa MWCO and spun at 4000g for 30-60 minutes. Once concentrated the nickel eluted protein was passed down a PD10 desalting column, which had been washed with 25mls of desalting buffer (buffer D). The initial elution was discarded but when a further 3.5mls of buffer D was passed through the PD10 column the elution was collected, as it contained imidazole free protein.

2.4.5 Size exclusion chromatography (SEC)

The PD10 elution was concentrated down to 0.5ml before being loaded onto a Superdex 200 Increase size exclusion column by injection, which had already been washed with two column volumes of buffer D. Typically 1.5 column volumes of 1ml fractions were collected running at a flow rate of 0.5ml/min. The AKTA FPLC provided a UV trace of the size exclusion column collection at A_{280}, allowing protein containing fractions to be identified.

2.4.6 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

The success of the protein purification was monitored by taking samples throughout and carrying out a sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis. Samples were mixed with an equal volume of 2x SDS loading buffer (100mM Tris-HCl pH6.8, 4% SDS, 0.2% bromophenol blue, 15% glycerol and 200mM Dithiothreitol) and boiled for ten minutes. These were then loaded onto a 12% SDS PAGE gel (where a 10ml separating solution contained: 3.35ml dH_2O, 2.5mls 1.5M Tris-HCl pH8.0, 4.0ml 30% acrylamide, 100µl 10% SDS, 50µl 10% APS, 15µl TEMED and a 5ml 4% stacking gel solution contained: 1.68ml dH_2O, 1.25ml 0.5M Tris-HCl pH6.8, 2mls 30% acrylamide, 50µl 10% SDS, 25µl 10% APS and 15µl TEMED). The SDS PAGE gel was loaded into a gel tank carried out in SDS-PAGE running buffer (25mM Tris-base, 192mM glycine and 0.1% SDS) at 200V for forty-five minutes. The gel was then stained by shaking gently in Coomassie stain for ten minutes ($1gL^{-1}$ Coomassie Brilliant Blue, methanol 50% v/v, glacial acetic acid 10% v/v and H_2O 40% v/v) which was then removed and replaced with a destaining solution (10% acetic acid, 30% methanol and 60% distilled water) and shaken gently for 1-2 hours.

2.4.7 Native polyacrylamide gel electrophoresis (Native-PAGE)

Precast NativePAGE Novex 4-16% Bis-Tris gradient gels were used to separate and monitor potential protein interactions and conformations in their native state. Protein samples were mixed with an equal volume of 2x native loading buffer (100mM Tris-HCl pH6.8, 02% bromophenol blue,

15% glycerol). Cold running buffer (25mM Tris-base, 192mM glycine) was added to the gel tank and samples loaded onto the precast gradient gels 4%-16%, purchased from BioRad, in which the gel tank had been placed in ice. The gel was run at a fixed current of 20mA for 3-6hours. The gel was then stained in Coomassise stain for ten minutes, which was then removed and replaced with a distaining solution and shaken gently for two hours.

2.4.8 Thrombin cleavage

A thrombin cleavage was used to remove the hexahistidine tag of the protein. First the protein of interest was buffer exchanged into PBS buffer (137mM NaCl, 10mM Na_2HPO_4, 2.7mM KCl pH 7.4) as described in section 2.4.3. Having measured the protein concentration 1 unit of thrombin was added per mg of protein. The cleavage was initially trialled by splitting the protein stock and thrombin mixture into 4 equal aliquots which were left shaking at 4^0C, 16^0C, 30^0C and 37^0C. Samples were taken after 1, 2, 3 and 16 hours and run on an SDS-PAGE to confirm the optimal temperature and time for successful thrombin cleavage.

2.5 Preparation of cobalamin intermediates

2.5.1 Lysis of cells

As the tetrapyrrole purification relies on the product trapping activity of a histidine tagged enzyme, an additional cell pellet which contained an individual Cob protein was also mixed with a tetrapyrrole containing pellet. (For example, 1-2 resuspended cell pellets of PC6B containing cells would also be mixed with one resuspended pellet of CobK containing cells). Each pellet was thawed on ice for thirty minutes and then resuspended in 30ml of buffer A, see Table 4. The resuspended cells were then lysed by sonication on ice for twelve minutes with thirty second intervals of 25-30 watts of sonication. The cell lysate was then centrifuged at 25,000g for thirty minutes at 4^0C and the supernatant was collected and stored on ice.

2.5.2 Purification of His-tagged protein

The supernatant was passed through a nickel column, which had been washed with five column volumes of buffer A. Once all the supernatant had been passed through the column, five column volumes of buffer A were washed through the nickel column. This was then followed by a second wash of five column volumes of buffer B. Finally, two column volumes of buffer C were added to the nickel column and collected as this elution contained the protein and cobalamin intermediate of interest as summarised in Figure 21.

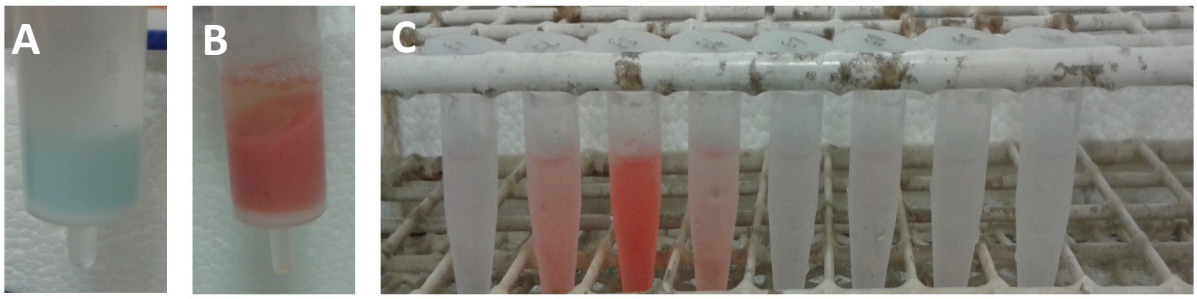


Figure 21: Example of a nickel column purification of the tetrapyrrole HBA where (A) is a pre-prepared nickel column (B) is a post-loaded nickel column and (C) shows 1ml aliquots of nickel column elution.

2.5.3 Protein buffer exchange

The nickel column elution was concentrated down to 2.5ml in a Vivaspin® 20 centrifugal concentrator with an appropriate molecular weight cut off (MWCO), depending on the protein being purified. This was typically a 10kDa MWCO or 30kDa MWCO, and spun at 4000g for 30-60 minutes. Once concentrated the nickel eluted protein and tetrapyrrole was passed down a PD10 desalting column, which had been washed with 25mls of buffer D. The initial elution was discarded but when a further 3.5mls of buffer D was passed through the PD10 column the elution was collected as it contained imidazole free protein and tetrapyrrole. At this stage if protein and cobalamin intermediate were required together then size exclusion chromatography was carried out as described in section 2.4.4 above.

2.5.4 Reverse phase chromatography

If the cobalamin intermediate was to be purified, then the pH of the PD10 elution was lowered to pH4 and any protein precipitation removed by centrifugation. The pH was then increased back to pH7.5 causing a second occurrence of protein precipitation which was removed by centrifugation. Finally, it was important to lower the pH again to pH4, otherwise the cobalamin intermediate would not bind to the reverse phase column in the next step. The pH4 cobalamin intermediate solution was then passed down a Telos® C18 reverse phase column. This column was initially washed with 5ml methanol, followed by 10ml ethanol and finally 5ml 1% acetic acid. The cobalamin intermediate was then passed through the column, which was then washed with 1.3ml 1% acetic acid. To elute the cobalamin intermediate 2ml of 50% ethanol was passed through the column and this elution was collected, Figure 22. This solution was then lyophilized for twelve hours and the powder of the cobalamin intermediate was stored at -80^{0}C until required.

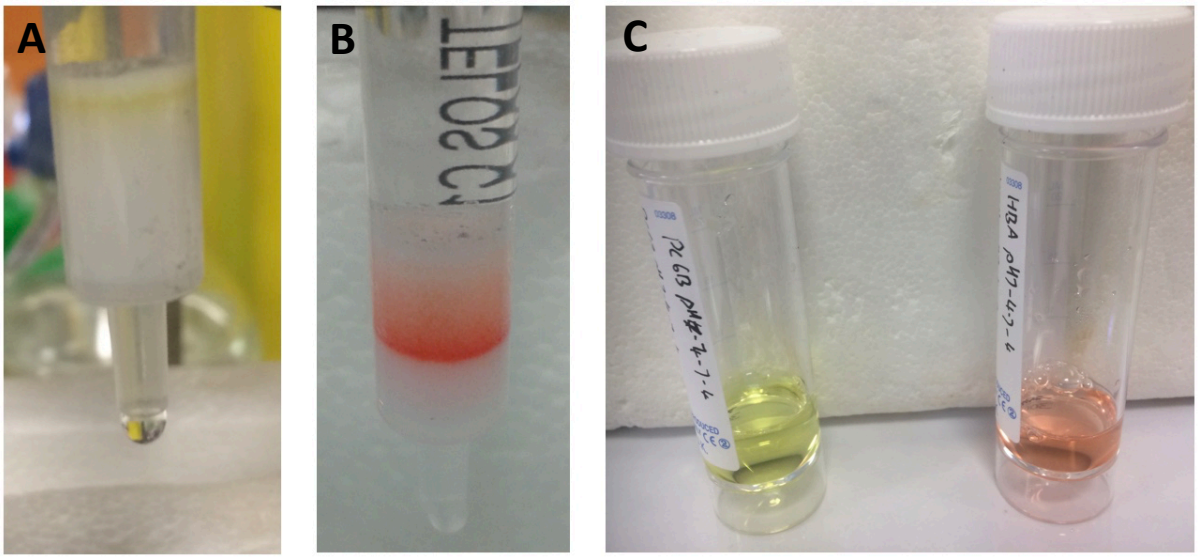


Figure 22: Example of a RP18 column purification of the tetrapyrroles PC6B (yellow) and HBA (pink) where (A) shows PC6B bound to the RP18 column (B) shows HBA bound to the RP18 column and (C) shows the elution collections of PC6B and HBA.

2.5.5 Mass spectrometry (MS)

The presence of the cobalamin intermediate could be monitored throughout both visibly, due to its colour, or by carrying out a UV-VIS wavelength scan between 300-500nm, where each tetrapyrrole gives a distinctive absorbance trace. The tetrapyrrole was confirmed by carrying out high resolution mass spectrometry and compared to the known molecular weight[53]. The mass spectrometry was kindly provided by Dr Roberto Buccafusca who is the Analytical Services Facility Manager at Queen Mary University.

2.6 Spectroscopy and biochemistry techniques

2.6.1 UV-VIS spectroscopy

UV-VIS spectroscopy was undertaken to monitor absorbance in both the ultraviolet (UV) region, 200-400nm, and the visible region, 400-800nm. The machine was zeroed before each run with an appropriate buffer solution and then the sample was loaded using a Quartz cuvette. Typically, a wavelength scan was performed to determine protein concentration or monitor the presence of tetrapyrroles.

Protein concentration was calculated using the Beer-Lambert Law, Equation 1, where A is the absorbance measured at 280nm, c is the protein concentration, ε is the protein's extinction coefficient and l is the length of the cuvette was 1cm. The protein's extinction coefficient was obtained by uploading the gene sequence into the ProtParam tool at https://web.expasy.org/protparam/.

Equation 1

$$A = c\,\varepsilon\,l$$

Where: A=Absorbance (AU), C= concentration (moldm^{-3}), ε=protein extinction coefficient (mol^{-1}dm^{3}cm^{-1}), l= length of cuvette (cm).

To follow the turnover reaction of tetrapyrrole enzyme modification, an appropriate wavelength was selected via an initial wavelength scan. The absorbance at this wavelength was then measured every three seconds for a set amount of time once the reaction had been initiated.

2.6.2 Enzyme kinetics using UV-VIS spectroscopy

The catalytic turnover of PC6B by the enzyme CobL was measured using UV-VIS spectroscopy. PC6B has an absorption maximum of 348nm when in the presence of CobK, whereas PC8 shows a decrease and shift with an absorbance maximum of 390nm. Due to the stronger absorption signal observed at 348nm this was chosen as the wavelength to measure absorption. PC6B concentrations were selected at 50μM intervals between concentrations of 50μM to 400μM. Initial studies were only carried out in the presence of CobL with a concentration of 5μM throughout. All reactions were initiated by adding 1mM SAM where absorption measurements were taken every three seconds for ten minutes. The second investigation of PC6B turnover was carried out in equimolar concentrations of CobK and NADPH to ensure that product binding occurred. A concentration of 5μM CobL was maintained and each run was initiated by the addition of 1mM SAM. All experiments were repeated three times to ensure that averages could be measured during the analysis which was carried out using GraphPad Prism version 7.0a for Mac OS X, GraphPad Software, La Jolla California USA, www.graphpad.com.

2.6.3 Fluorescence spectroscopy

Fluorescence spectroscopy was used to observe the presence of a fluorophore such as HBA and typically relies on a molecule having conjugated π-electrons present in the structure. This technique relies on a particular wavelength to excite electrons from a ground state to a higher energy level, which upon returning to the ground state emit light radiation, which is measured at a perpendicular angle to the beamline. An appropriate excitation wavelength was selected based on the UV-VIS

wavelength scan of the fluorophore to identify where maximum absorbance/excitation occurs. Detection was measured using a wavelength scan at higher wavelengths to the selected excitation wavelength.

2.6.4 Real-time PCR thermal melts

A real-time PCR instrument was used to screen the thermal shift profiles of proteins with and without ligands in a 96-well format. Temperature profiles were generated by averaging results measured across three repeats. The well conditions were always made to a final volume of 20µl upon which a 1x solution of the dye SYPRO® Orange was added. The 96 well tray was sealed with SealPlate® film and the thermal shift assay was carried out at 20-95^{0}C, where the temperature was increased by 1^{0}C every minute.

2.6.5 Isothermal calorimetry (ITC)

Isothermal calorimetry is a sensitive technique that is able to measure the heat that is either released or absorbed during a protein binding event. As CobK protein was less precious and easier to produce than PC6B, the protein was added to the cell of the ITC machine while PC6B was added to the syringe, as the volume required for the cell was 280µl compared to 60µl in the syringe. Initially a 10x concentration of ligand: protein was trialled using CobK with a concentration of 30µM and a PC6B concentration of 300µM. Initial runs were trialled where PC6B was injected across 13 injections of 3µl into CobK in the presence and absence of NADPH (30µM). As ligand saturation occurred too quickly the number of injections was increased to 19 with a volume of 2µl and the concentration of protein was increased to 45µM to increase the signal and the temperature lowered to 15^{0}C given the reaction observed was endothermic.

2.7 Structural biology

2.7.1 X-ray crystallography background and methodology

2.7.1.1 Introduction to X-ray crystallography

X-ray crystallography is the most powerful tool to determine the atomic resolution structure of proteins of any size, provided that high-quality crystals and diffraction data can be collected. The crystallography pipeline is summarised in Figure 23A, highlighting the key stages involved in solving a protein structure. Initially this is begun by cloning the protein of interest into a vector, expressing the protein in high amounts and subsequently purifying said protein, as described above in Section 2.4.

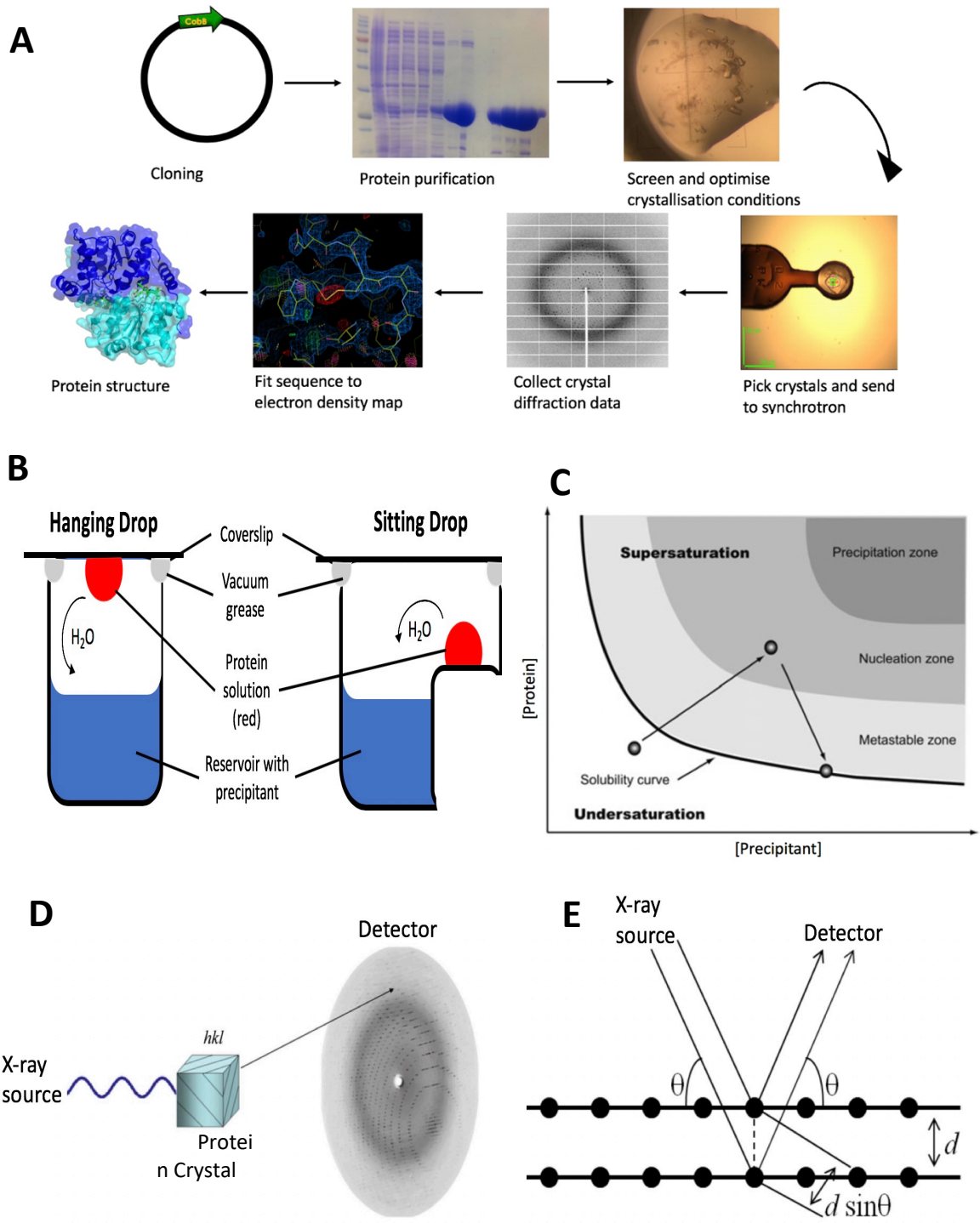


Figure 23 A) Show an overview of the stages involved in the crystallography pipeline when trying to solve a structure. B) Shows the two methods used to grow and produce crystals using either the hanging drop or sitting drop method, upon which a sample of protein solution is mixed with reservoir solution and left so that water vapour can evaporate and concentrate the protein. C) Shows a diagram of how protein solubility is linked to concentration of the protein and precipitant within the solution. D) Shows a simplified diagram of how diffraction is collected from a crystal at a synchrotron, where X-rays are shone onto the crystal at different angles and diffraction data collected on a detector, figure taken from Taylor 2010 [85]. E) Shows the relationship between Bragg's equation to determine the distance d, between the lattice planes of a crystal during a diffraction data collection.

2.7.1.2 X-ray crystallography screening and optimisation

Once pure protein was obtained it was concentrated to 10-20mg/ml and used to screen for successful crystal conditions. Screening was carried out using 96-well plates using either the hanging drop or sitting drop method as shown in Figure 23B. For protein crystals to form the protein needs to initially be undersaturated and soluble in a screening solution. Figure 23B and C show that as the reservoir of the screening solution draws water from the protein drop, this leads to an increase in the concentration of the protein and precipitant in the protein drop, allowing nucleation of protein crystals to form[86]. If nucleation is successful, the concentration of the protein in solution will decrease and crystal growth favoured if the solubility lies within the metastable zone.

Crystal trays were prepared by trialling crystal screening solutions which were purchased from Molecular Dimensions Limited and included; Structure Screen 1+2 HT-96 (MD1-30), JCSG-plus™ HT-96 (MD1-40), MemGold 2™ (MD1-63), PACT *premier*™ (HT-96) and Morpheus® HT-96 (MD1-47). These trays were prepared by mixing pure protein with crystal screening solutions using ttplabtech mosquito®. Trays were sealed and left at 19^0C and were checked regularly for two weeks.

If any screen conditions looked promising these were optimised. This can involve altering physical, chemical and biochemical factors[87]. Typically, chemical factors were changed by altering the salt or precipitant concentration across a small range. This was undertaken using a Celltrat 24 Well Cell Culture Plate, where 1µl of crystal reservoir solution was mixed with 1µl protein solution, placed on a cover slip and sealed above the respective reservoir with vacuum grease using the hanging drop method. These trays were again left at 19^0C and checked regularly for two to three weeks.

2.7.1.3 Picking crystals

Protein crystals were picked using appropriately sized Mounted Round LithoLoops down a microscope. Unless the Morpheus screen condition had been used, the protein crystals were then transferred into a cryo-solution which contained the crystallisation solution and 10-20% glycerol. The loops were then transferred into a Uni-Puck which was cooled under liquid nitrogen. Once all

the crystal samples had been loaded, the Uni-Puck was closed and transferred to a Dewar that had been cooled using liquid nitrogen. Through a collaboration with Birkbeck University the Dewar was shipped off to Diamond Light Source synchrotron based in Oxfordshire ready for data collection. Crystal samples were uploaded on to https://ispyb.diamond.ac.uk and crystal diffraction data was usually collected remotely by logging into the Diamond beamline via the software NoMachine.

2.7.1.4 Collecting crystal diffraction data

Diffraction data of protein crystals are obtained by illuminating with X-rays, where a single oscillation image is obtained by rotating the crystal through a small degree of rotation, Figure 23D[85,88]. These images are combined to form a data set that samples the entire three-dimensional diffraction pattern for said sample. Interactions of X-rays with crystals are described using the Bragg's model, in which the diffraction images are a result of reflections within the lattice planes of the crystal[89]. Every atom within the repeating unit of a protein crystal contributes to the intensity of every reflection in a diffraction pattern and these reflections are related to the planes from which they have arisen by the use of the crystal lattice indices *h,k* and *l*. Bragg's law explains that two beams with identical wavelength and phase are scattered off two different atoms within it, Figure 23E. The lower beam in Figure 23E travels an extra length of $2d\sin\theta$ and constructive interference occurs when this length is equal to whole number integer of the wavelength giving rise to Bragg's equation, Equation 2, thus allowing the amplitude to be measured on the detector via the intensity of the diffraction spots [88,89]. This equation shows the relationship between the interplanar distances of the reflecting planes d_{hkl} and the angle of reflection θ_{hkl} is reciprocal, where for example closely spaced planes give rise to large diffraction angles[89].

Equation 2

$$n\lambda = 2d_{hkl}sin\theta_{hkl}$$

Bragg's equation where λ is the wavelength of the X-ray radiation, n is a whole number integer, d is the spacing of the interatomic planes between the lattice planes of the crystal *hkl* and θ the angle of reflection.

2.7.1.5 Calculating the electron density map

In order to generate an electron density map of the structure both the amplitude of the diffracted X-ray waves and their phase angles are required. During a diffraction experiment the intensities of reflected waves from the lattice planes *hkl* from the crystal are measured via a detector, Figure 23D. The amplitude of the reflected wave $|F_{hkl}|$ is proportional to the square root of the measured intensity. Within the unit cell of the crystal the electron density at a given position (xyz) can be calculated using Fourier transformation shown in Equation 3, in which the electron density at (xyz) is the sum of contributions to the point (xyz) of waves scattered from the crystal planes (*hkl*), where the amplitude relies on the number of electrons in the plane, added with the correct relative phase of the wave (a_{hkl}). The amplitudes are measured during diffraction of the crystal however the phases are lost during the experiment and this is typically coined the phase problem[88,90]. When phase information is available a Fourier transformation can be used to construct an electron density map in reciprocal space. Phases can be obtained either by using the structure of a homologous protein in a process known as molecular replacement or via the incorporation of a heavy atom[88].

Equation 3

$$p(xyz) = \frac{1}{V} \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i(hx + ky + lz))$$

Fourier transformation equation used in X-ray crystallography to calculate the coordinates of electron density, p(xyz), in a given unit cell with volume (V). *hkl* represent the lattice planes of the crystal, $|F_{hkl}|$ is the structure-factor amplitude of the X-ray beam and α_{hkl} is the phase associated with the amplitude[90].

To determine a crystal diffraction collection strategy, crystals were loaded and centred in the synchrotron beamline and three images were taken at a 45^0 oscillation. A successful crystal screen would result in an appropriate strategy being generated and applied, which allowed for a thorough diffraction data collection. Crystal diffraction images were processed using Dials or Mosflm, via the CCP4 Suite[91]. Protein phases were determined and refined using molecular replacement through Phaser via the Phenix suite[92].

To solve the phases molecular replacement was used and relies on a Patterson function, where a Patterson map can be calculated from the same Fourier summation that is used to calculate the electron density map[85]. A Patterson map is also produced from the atomic coordinates of a structurally similar model and rotated over the calculated Patterson map obtained from the structure-factor amplitudes of the protein crystal structure to determine the orientation of the model in a new unit cell[85] . A Patterson function is a Fourier summation that can be calculated directly from the measured intensities [93]. The translation of the correctly oriented model can also be measured via a translational function z-score (number of standard deviations above the mean value), in which a score above 8.0 means the structure of interest can definitely be solved.

Phases can also be solved experimentally by incorporation of heavy atoms into the crystal lattice[88]. This method relies on anomalous scattering of X-rays which give rise to measurable

intensity changes which allow the position of the heavy atom to be determined[85]. This method was attempted to solve the structure of CobB exploiting the incorporation of selenomethionine (SeMet) residues. To obtain phase information from the incorporation of selenium multiple anomalous diffraction (MAD) phasing was carried out. This technique requires the presence of a heavy atom which has an X-ray absorption edge in the energy range of the X-ray wavelength used (~1Å) [93]. Three different wavelengths were applied around the heavy atom's absorption edge and the change in wavelength causes a change in the measured intensities of the individual reflections allowing these differences to be used to directly calculate the phase angle for each reflection [93].

This was undertaken using the SelenoMet™ Medium Complete (MD12-500) purchased from Molecular Dimensions Limited. A 100ml LB culture was inoculated with a single cell colony and grown overnight for 12-16 hours at 37^0C with the appropriate antibiotics. The cells were centrifuged at 4000g for fifteen minutes and the pellets washed three times with sterile water. The pellets were resuspended in 1ml of sterile water and inoculated into a litre of prewarmed (30^0C) minimal media which contained selenomethionine. These cultures were grown for two hours at 37^0C at 220rpm before the addition of 1mM IPTG was added to induce protein production, at which stage the temperature was lowered to 18^0C and then left to grow for a further 12-16 hours. Cells were harvested by centrifugation at 8,000g for fifteen minutes and stored at -20^0C. The protein purification was identical to section 2.4 described above.

The selenomethionine medium was prepared by dissolving 21.6g of SelenoMet Medium Base in one litre of dH_2O and autoclaved. Subsequently 5.1g of SelenoMet Nutrient mix was dissolved in 50ml dH_2O and filter sterilised before being added to the autoclaved SelenoMet Medium Base solution. A further 4ml of 250x selenomethionine and appropriate antibiotics were added before the grow up.

2.7.1.6 Structure validation

Initial estimates of phases typically have large errors which makes the electron density map noisy and ill defined[88]. Therefore, structure refinement and validation are carried out. This relies on model refinement programs and graphic displays such as COOT and Molprobity to help provide validation to interpret many unexplained difference peaks in the electron density Fourier map [94,95]. Furthermore calculated phases are also prone to model bias and these errors can be measured using R-factors and R-free [96,97]. The R-free is a useful validation as it is calculated by removing a random sample of reflections and used to compare to the refinement R-factors to provide a more objective quality criterion as standard R values can easily be abused [97,98]. Additionally cycles of refinement involve manual rebuilding of the model until a good agreement can be seen between the model and the calculated electron density map and these cycles include further validation of stereochemical quality indicators, such as acceptable deviations in bond length, amino acid rotamer orientations and torsion angles via the Ramachandran plot [95,98].

2.7.2 Nuclear magnetic resonance (NMR)

2.7.2.1 N^{15} labelled media

Isotopically labelled ^{15}N proteins were produced by growing cells in minimal media in which ^{15}N labelled NH_4Cl replaced regular ammonium chloride. A method described by Marley et al[99] was followed as it argued that higher protein yields could be obtained by initially growing cells in unlabelled rich media to produce all the cell machinery required for protein production. Once the cells were in the exponential growth phase they were transferred into ^{15}N labelled minimal media and protein production was induced.

This was achieved by initially growing cells to an optical density of OD_{600} ~0.7 in one litre batches of LB media as described above in section 2.3.3. These cells were then pelleted by centrifugation at 5000g for thirty minutes. The cells were washed in M9 salt solution (Table 3) and then resuspended

in isotopically labelled minimal media and inoculated into 1L batches of ^{15}N labelled minimal media. The cells were grown for a further hour at 37^0C so as to allow for growth recovery and any unlabelled metabolites to be cleared, before protein expression was induced by the addition of 0.8mM IPTG. The temperature was then lowered to 18^0C at 220rpm and the cells left to grow for a further eighteen hours. Cells were harvested by centrifugation at 8000g for fifteen minutes and cell pellets stored at -20^0C until required. N^{15} labelled protein was purified as normal as described in section 2.4 above.

2.7.2.2 NMR data collection

Pure ^{15}N labelled protein was used for 2D HSQC experiments on the Bruker 700MHz NMR machine at the Francis Crick Institute. The data collection and results output were facilitated under the supervision of Dr Shuang Gu, from Queen Mary of London University. Protein concentrations no lower than 0.2mM were used and protein was kept in buffer D solution throughout.

2.8 Bioinformatics

Docking data of precorrin intermediates and products had been carried out by members of the Pickersgill group before this project was undertaken, as a number of methyltransferase crystal structures had been solved lacking precorrin ligands. Multiple models of different conformations in which precorrin substrates/product bind to their respective methyltransferase were generated using AutoDock Vina webserver[100]. In some cases the cleft opening was too tight on currently solved structures so certain amino acids were made more flexible using the elNemo web server for normal mode analysis to generate new protein data bank templates for docking[101]. Some structures of the ligands were energy-minimized by Avogadro: an open-source molecular builder and visualization tool. Version 1.XX. http://avogadro.cc/ [102].

Chapter 3: Crystallography of Cob proteins

3.1 Introduction, aims and objectives

The Cob proteins produced in the laboratory during this project included CobK, CobL, CobL^{c} (C-terminal domain of CobL), CobH and CobB. Three of these have had their protein structure solved using X-ray crystallography: CobK, CobL^{c} and CobH[53,62,67]. The crystal structures of CobK and CobH provide evidence of product binding activity based on the sequestration of their respective products PC6B and HBA. This unusual behaviour was first observed visually in the laboratory due to these brightly coloured cobalamin intermediates accompanying the enzymes during purification as shown in Chapter 2. Based on this product trapping by CobK and CobH these proteins can be grouped into the pairs CobK-CobL and CobH-CobB as the second enzyme in the pair must obtain its substrate for the cobalamin biosynthetic pathway to proceed. As the enzymes CobL and CobB are both proteins larger than 30kDa, X-ray crystallography studies would be able to determine their structures. Therefore, to explore these pairs further and help understand how they work and provide guidance on future mutagenesis studies, the aims were to:

1. Solve the structure of full length CobL using X-ray crystallography.
2. Solve the structure of CobB using X-ray crystallography.
3. Co-crystallise and solve the structure of the enzyme pairs CobK-CobL and CobH-CobB using X-ray crystallography.

3.2 CobL crystallography

3.2.1 CobL protein purification

The first step in attempting to solve the crystal structure of CobL was to purify the protein. The purification of CobL became a fairly routine procedure, as all the Cob proteins were purified in a highly similar fashion using a nickel column followed by size exclusion chromatography. However, it was found that CobL was far more prone to insolubility during purification and therefore the buffers used required a much higher sodium chloride concentration of 500mM compared to the typical 100mM for the other Cob proteins to maintain solubility. Despite this small set back the purification of CobL did produce pure protein as seen in Figure 24. Once pure protein was obtained it was concentrated typically to concentrations between 10-20mg/ml ready for crystallography.

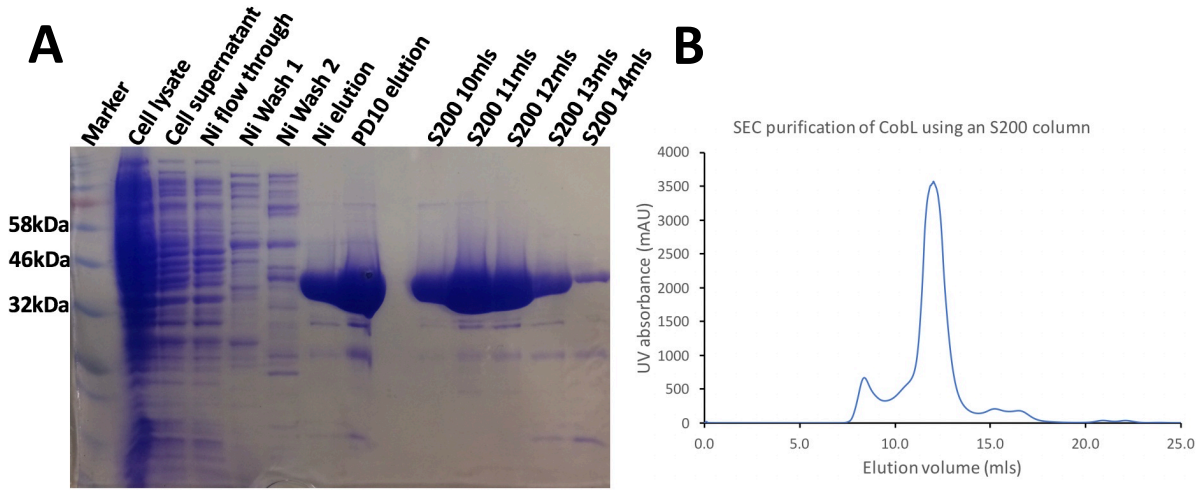


Figure 24: Purification of CobL (~42kDa). A) SDS-PAGE analysis of the protein purification process. B) UV profile of the S200 size exclusion column elution where the peak is from the fractions 11 to 14mls.

3.2.2 CobL crystal screening and optimisation

A thermal melt study carried out on a real time PCR machine of 150μM CobL with an excess of 20x substrates PC6B, SAH and SAM was carried out a real-time PCR machine and shown in Figure 25. When CobL was in the presence of one ligand only Figure 25A, the addition of SAM caused an increase in melting temperature (T_m) from 38^0C to 49^0C while the addition of SAH had an even greater effect causing an increase in T_m from 38^0C to 56^0C, while PC6B had no effect on T_m. The addition of two ligands, Figure 25B, resulted in the same respective T_m when one ligand was used demonstrating that PC6B has no overall effect on the increased stability of CobL and this increase in T_m is solely due to SAH and SAM binding only.

Therefore, for CobL crystal screening a 20x times excess of SAH and SAM was added separately to CobL samples before crystal trays were set up. The Molecular Dimensions screening conditions 'Structure Screen I & II' (SSI/II) and 'JCSG+' were trialled and after a few days small crystals were observed as summarised in Figure 26. From these initial screens the conditions that yielded the best looking crystals were selected for optimisation. These were SSI-C8, SSI-C10 and SSII-E12. Crystals were picked and soaked in a cryogenic solution containing the addition of 20% glycerol where they were sent off to the Diamond Synchrotron.

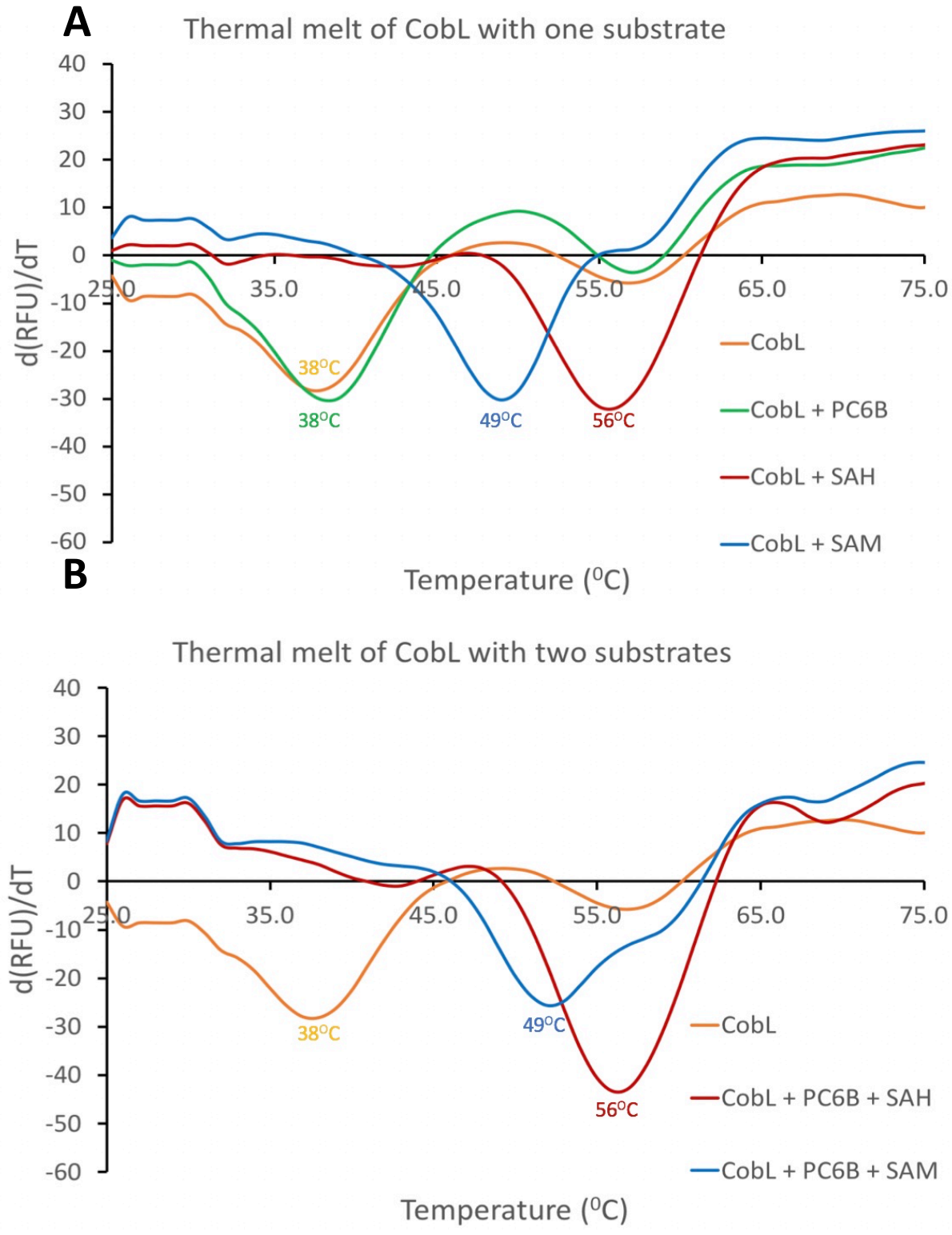
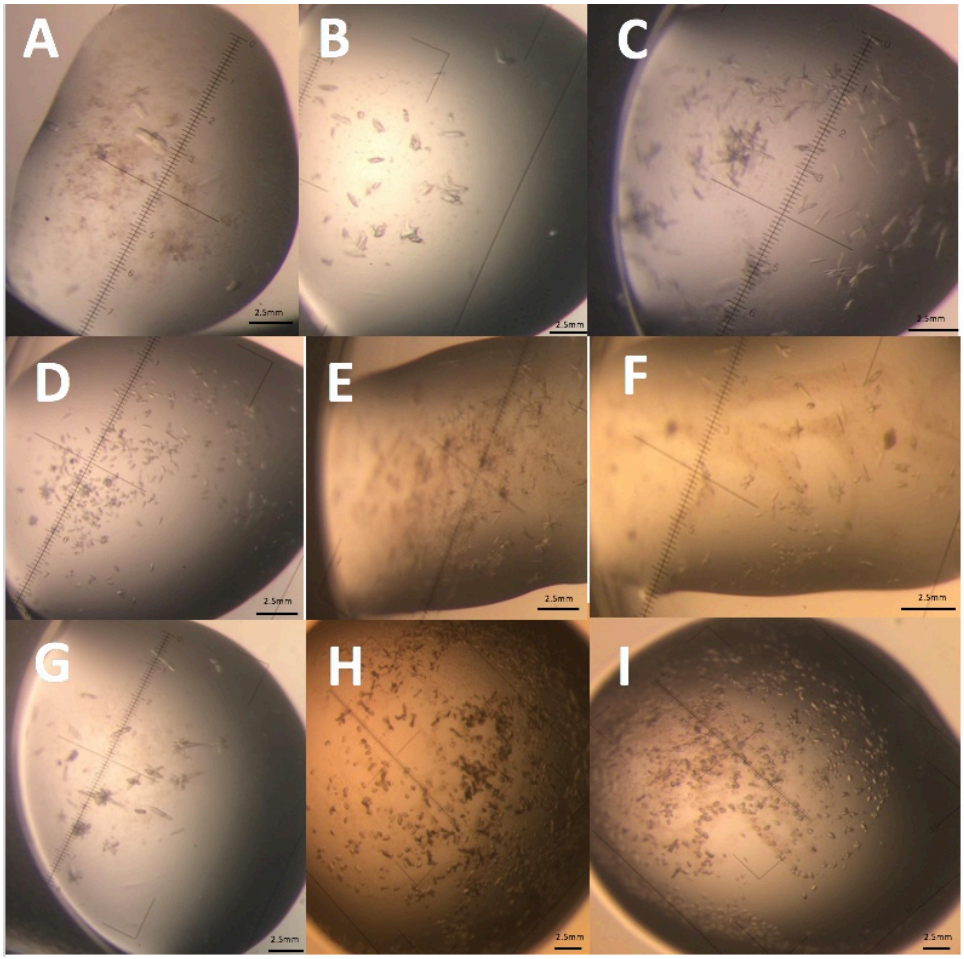


Figure 25: A real time PCR thermal melt study of CobL in the presence of one (A) or two (B) ligands which included PC6B, SAM and SAH. The melting temperature (T_m) can be calculated by recording the temperature at which the negative first derivative occurs and has been labelled on each curve.



Panel	Crystal screen	Crystal Conditions
A	SSI-C3	1.5M lithium sulphate, 0.1M sodium HEPES pH7.5
B	SSI-C8	2.0M ammonium sulphate, 0.1M Tris-HCl 8.5
C	SSI-C10	0.2M sodium citrate tribasic dihydrate, 0.1M Tris-HCl pH8.5, 30% v/v PEG 400
D	SSII-E10	0.01M nickel (II) chloride hexahydrate, 0.1M Tris-HCl pH8.5, 20% w/v PEG 2000 MME
E	SSII-G5	2.0M ammonium sulphate, 0.2M potassium sodium tartrate tetrahydrate, 0.1M sodium citrate pH5.6
F	SSII-G10	0.01M manganese (II) chloride tetrahydrate, 0.1M sodium citrate pH5.6, 2.5M 1,6-hexane diol
G	SSII-E12	0.1M sodium HEPES pH7.5, 10% w/v PEG 6000, 5% MPD
H	SSII-E12	Optimization of SSII-E12 where PEG 6000 was varied 9-15% w/v
I	SSII-E12	Optimization of SSII-E12 where PEG 6000 was varied 9-15% w/v

Figure 26: CobL crystal screening and optimisation with crystallography conditions summarised below. Structure Screen I (SSI) and Structure Screen II (SSII) were used.

3.2.3 CobL crystal data processing

Data were collected at Diamond Light Source using an X-ray wavelength of 0.97628Å on beamline I04. The highest resolution reflection observation was 3.6 Å. These images were integrated using iMosflm and scaled using Aimless through the CCP42i Suite. The crystallographic statistics of the data processing are summarised in Table 5. The most probable space group was $P4_2$ 2 2 based on the R_{pim} and trials of numerous space groups. The point group summary had a space group confidence score of 0.711, with a Laue group confidence and probability of 0.989 and 0.992 respectively. The data continued to be confirmed as highly twinned, even when analysed in space group P1, Figure 27. Molecular replacement was attempted using phaser via the Phenix Suite. Models which were used included $CobL^c$, CbiET and as the N-terminus does not have a structure a CobL sequence identical model based on CbiE was generated via the phyre2 server, which is a protein fold recognition server. However molecular replacement solutions were not convincing with the best TFZ score of 5.1. So, the search models could not be placed in the unit cell satisfactorily and molecular replacement was therefore unsuccessful.

CobL data collection statistics

Space group	P 4_2 2 2		
Cell edges (Å)	141.750	141.750	114.880
	90.000	90.000	90.000
Wavelength (Å)	0.97628		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	3.60-89.25	8.82-89.25	3.60-3.94
Number of observations	102,591	6944	23,638
Number of unique reflections	13,990	1039	369
Multiplicity	7.3	6.9	7.2
Rmerge	0.239	0.078	0.681
Rpim	0.087	0.031	0.251
Rmeas	0.256	0.085	0.729
Mean I/σ	5.7	13.3	2.6
$CC_{1/2}$	0.99	0.993	0.865
Completeness (%)	98.7	97.4	98.9
Wilson B estimate ($Å^2$)	64.21		

Table 5: Summary of data collection statistics for CobL crystals where Rmerge = Σ hkl Σ i |Ii–< I>| /– Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge[62,103].

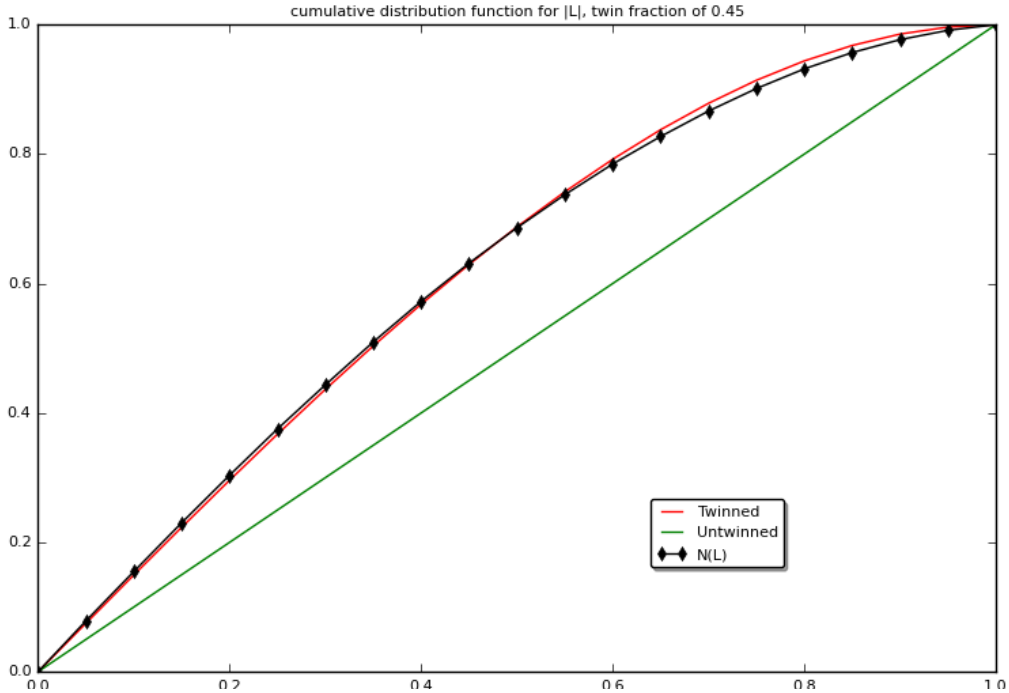


Figure 27: L-test showing that the CobL crystals were highly twinned, with a twin fraction of 0.45, where the CobL data (black line) is almost completely overlapping with the twinned function (red line). Graph generated by CCP42i.

3.2.4 CobL co-crystallisation with CobK

A co-crystallisation of CobL with CobK was attempted in the hope that if these proteins interact, they may be trapped and observed in a crystal structure. CobK was purified as described in Chapter 2 section 2.4 and a typical purification profile is shown in Figure 28. Once pure CobK and CobL proteins were obtained they were mixed in equal molar concentrations of 20mg/ml with a 10x excess of substrates PC6B, NADPH and SAH present to encourage product trapping by CobK and prevent PC6B turnover by CobL.

Protein crystals were observed, Figure 29, in a screen tray of Structure Screen II-F4, which was composed of 1M sodium acetate trihydrate, 0.1M sodium HEPES pH7.5 and 0.05M cadmium sulphate $^{8}/_{3}$-hydrate. These crystals were picked straight from the sitting drop of the screen tray and sent off to collect diffraction data at the Diamond Synchrotron. A summary of the crystal data statistics are reported in Table 6, where the diffraction images were integrated using Dials in the CCP4-2i suite followed by molecular replacement using Phaser in the Phenix suite.

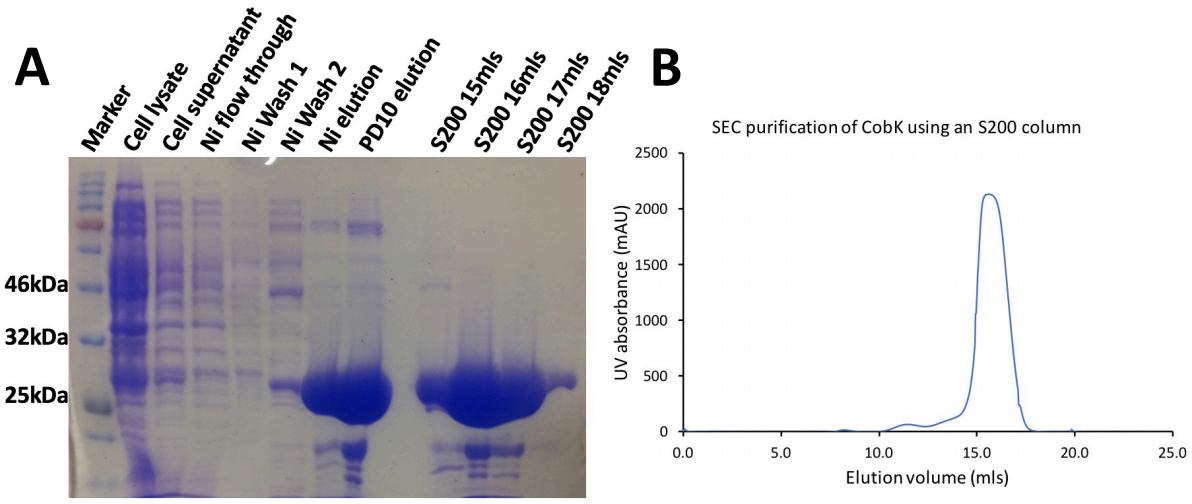


Figure 28: Purification of CobK (~26kDa). A) SDS-PAGE analysis of the protein purification process. B) UV profile of the S200 size exclusion column elution where the peak is from the fractions 15 to 18mls.

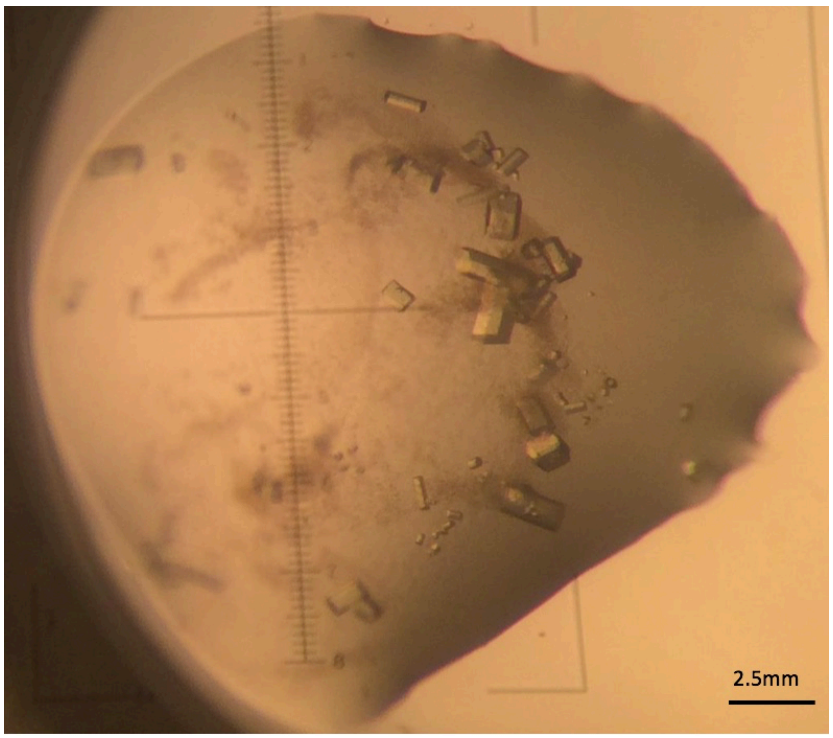


Figure 29: Crystals obtained from screening attempts of CobK mixed with CobL using Structure Screen II containing 1M sodium acetate trihydrate, 0.1M sodium HEPES pH7.5 and 0.05M cadmium sulphate $^{8}/_{3}$-hydrate.

CobK data collection statistics

Space group	P 1 2_1 1		
Cell edges (Å)	33.645	113.113	58.340
	90.000	93.293	90.000
Wavelength (Å)	0.91587		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	2.00-58.24	8.94-58.24	2.00-2.05
Number of observations	148,346	1873	10,984
Number of unique reflections	29,413	351	2,162
Multiplicity	5.0	5.3	5.1
Rmerge	0.114	0.059	0.75
Rpim	0.056	0.028	0.408
Rmeas	0.127	0.066	0.408
Mean I/σ	7.7	16.9	2.1
$CC_{1/2}$	0.995	0.998	0.663
Completeness (%)	100	99.5	100
Wilson B estimate ($Å^2$)	28.15		

Refinement statistics

Resolution (Å)	2.00-58.24
Number of reflections	25100
R-work/R-free	0.24/0.31
Number of atoms (protein chains/ligand/water)	2/0/284
Rmsd bond length	0.08
Rmsd bond angle	0.923
Mol Probability	1.87
Residues in most favoured regions (%)	96.65
Outliers (%)	0

Table 6: Data collection and refinement statistics for CobK and CobL co-crystallisation, which upon using molecular replacement the unit cell only contained CobK. Rmerge = Σ hkl Σ i |Ii–< I>| /− Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge R-factor = Σ hkl |Fo–Fc |/Σ hkl Fo, where Fo and Fc represent the observed and calculated structure factors, respectively. The R-Factor is calculated using 95% of the data included in refinement and R-free the 5% excluded[62,103].

Upon solving the diffraction data using molecular replacement, it became apparent that the protein crystals only contained the protein CobK. The crystals were obtained in a different crystallisation condition to the already solved structure and was in a different space group, with much tighter crystal packing in the unit cell. No ligands were present in the CobK structure meaning a higher resolution structure of the apoenzyme was obtained to 2.00Å. This was a significant improvement compared to the current structure of the apoenzyme which has a resolution of 3.17Å.

Upon refining the crystal structure three features were observed. Firstly, the disulphide bond present in the known published structure was not actually present, Figure 30. Secondly a metal ion was observed in the structure of both protein chains due to the octahedral geometry between the R-groups of threonine 72 and serine 80, the backbone of proline 74 and alanine 76 and two additional water molecules, located near the NADPH binding site, as shown in Figure 31 where the sodium-oxygen distances are too short for this to be a water molecule. Finally in chain A, one of the two protein chains present in the unit cell, it was possible to partially model the loop region of the apoenzyme, which has not been observed in the current published structure of the apoenzyme, Figure 32. The reason the dynamic loop region was observed and could be modelled was due to the tight crystal packing present which resulted in crystal contacts between separate protein chains which somewhat restrained the flexible loop region to some degree. The two protein chains in this data collection were back to back, where one was inverted, and a large number of water molecules bridge the protein-protein interface.

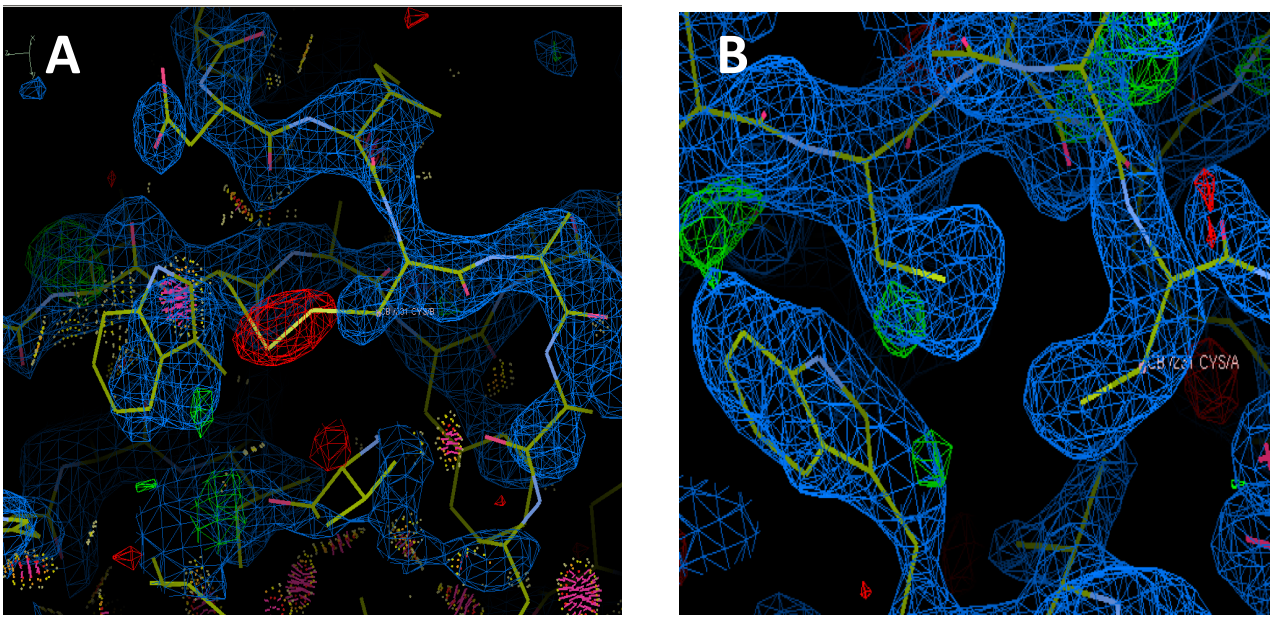


Figure 30: Electron density map of CobK showing that the disulphide bond does not occur as highlighted in red during pre- and post-refinement in (A) and (B) respectively. Figure produced via COOT[94].

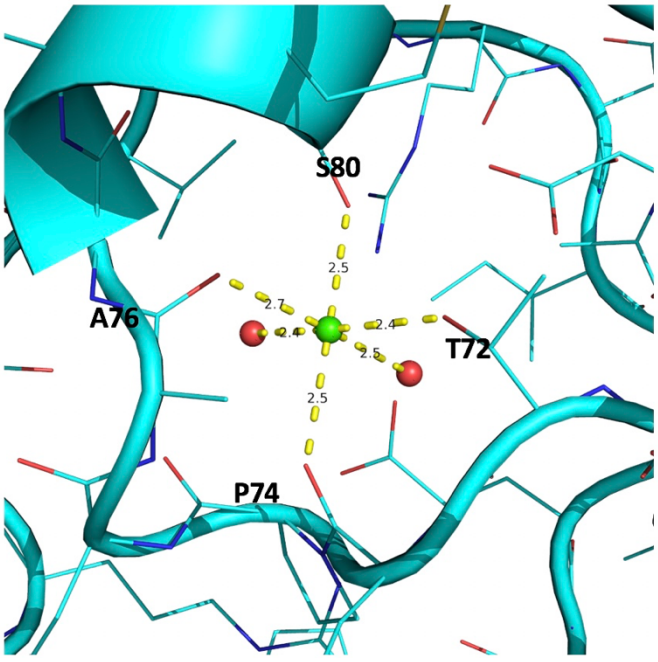


Figure 31: Octahedral coordination of a sodium ion (green) binding in CobK crystal data, which was an artefact of the crystallisation conditions where 1M of sodium acetate was also present. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

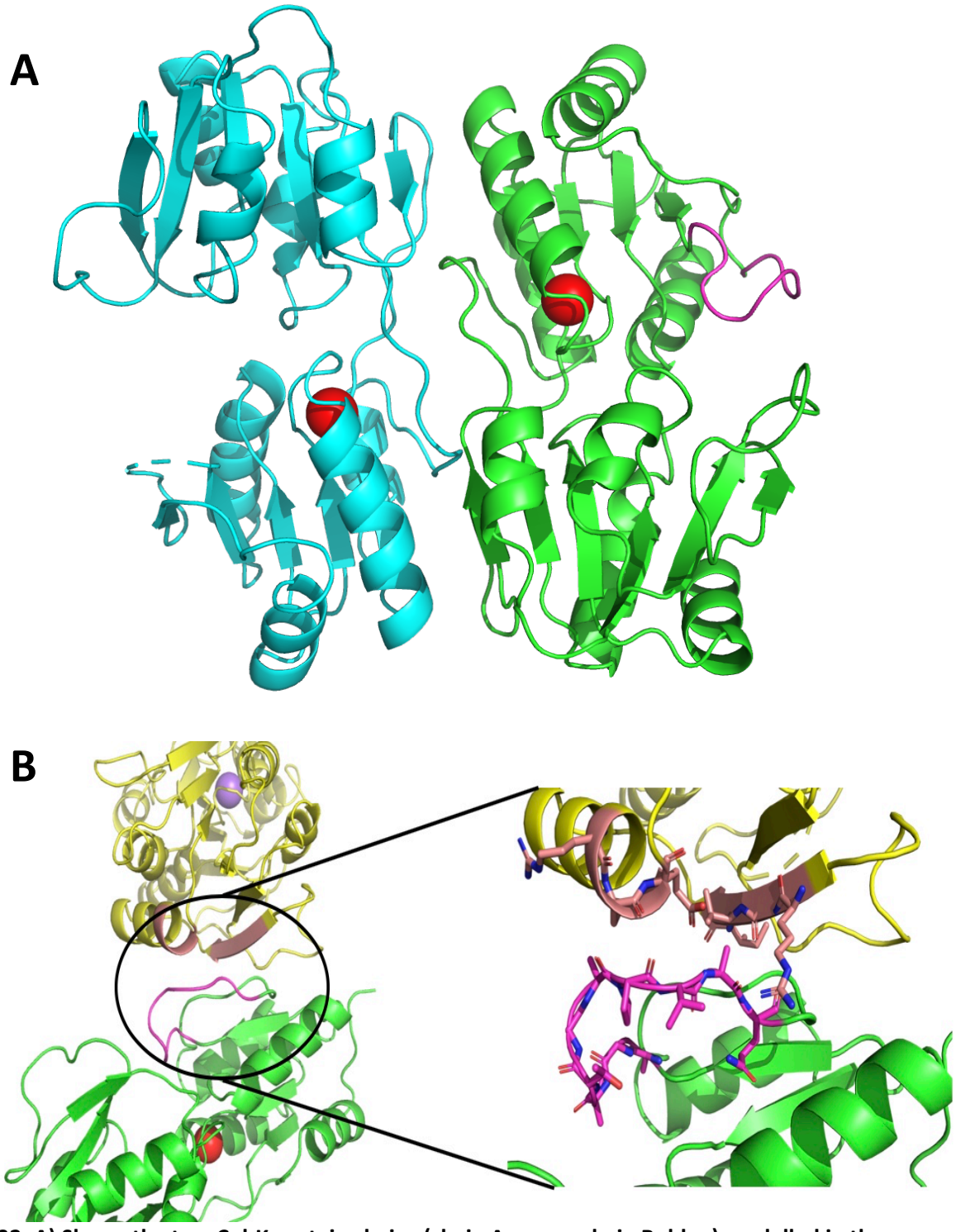


Figure 32: A) Shows the two CobK protein chains (chain A-green, chain B- blue) modelled in the asymmetric unit, with each containing a sodium ion coloured in red. Chain A has the flexible loop region highlighted in purple, while this loop could not be modelled in chain B, due to its flexible nature. B) Shows the close proximity of a neighbouring protein chain (yellow) which due to crystal contacts (pink) allowed this flexible loop region (purple) to be modelled. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

3.3 CobB crystallography

3.3.1 CobB protein purification

Initially purification of CobB typically resulted in low yields of protein which would subsequently degrade over a 24-hour period. To resolve this a number of steps were introduced. This included keeping all buffers on ice throughout the purification. The addition of 2mM TCEP to the purification buffers was included as the protein contained an active site cysteine, in which TCEP would prevent its oxidation. While these changes did result in slightly better yields it was concluded that purifying the protein in the presence of ligands would assist in stability and lead to improved yields. To achieve this CobB was purified from cell pellets obtained from HBAD tetrapyrrole producing cells as described in section 2.3.4 and this did result in greater protein yields as shown in Figure 33. Furthermore, as these cells were grown over a 24-hour period it was concluded that there was probably more HBAD present than protein. Therefore, an additional cell pellet containing histidine tagged CobB was also mixed into the purification during the sonication step so that any free HBA/HBAD present in the cell lysate could be mopped up and assist in increased protein yields. CobB could visibly be observed binding HBAD during the purification protocol exhibiting product binding activity similar to CobH as shown in Chapter 2 section 2.5.2 Figure 21.

To determine if the presence of substrates assisted in the protein stability of CobB thermal melt studies were carried out using a real time PCR instrument. These are summarised in Figure 34 in which one, two or three substrates were present in an excess of 20x to CobB. Each experiment was repeated three times and an average plotted to determine the melting temperature (T_m). Native CobB gave a T_m of 58°C and in the presence of one, two or three substrates the measured T_m remained the same as shown in Figure 34.

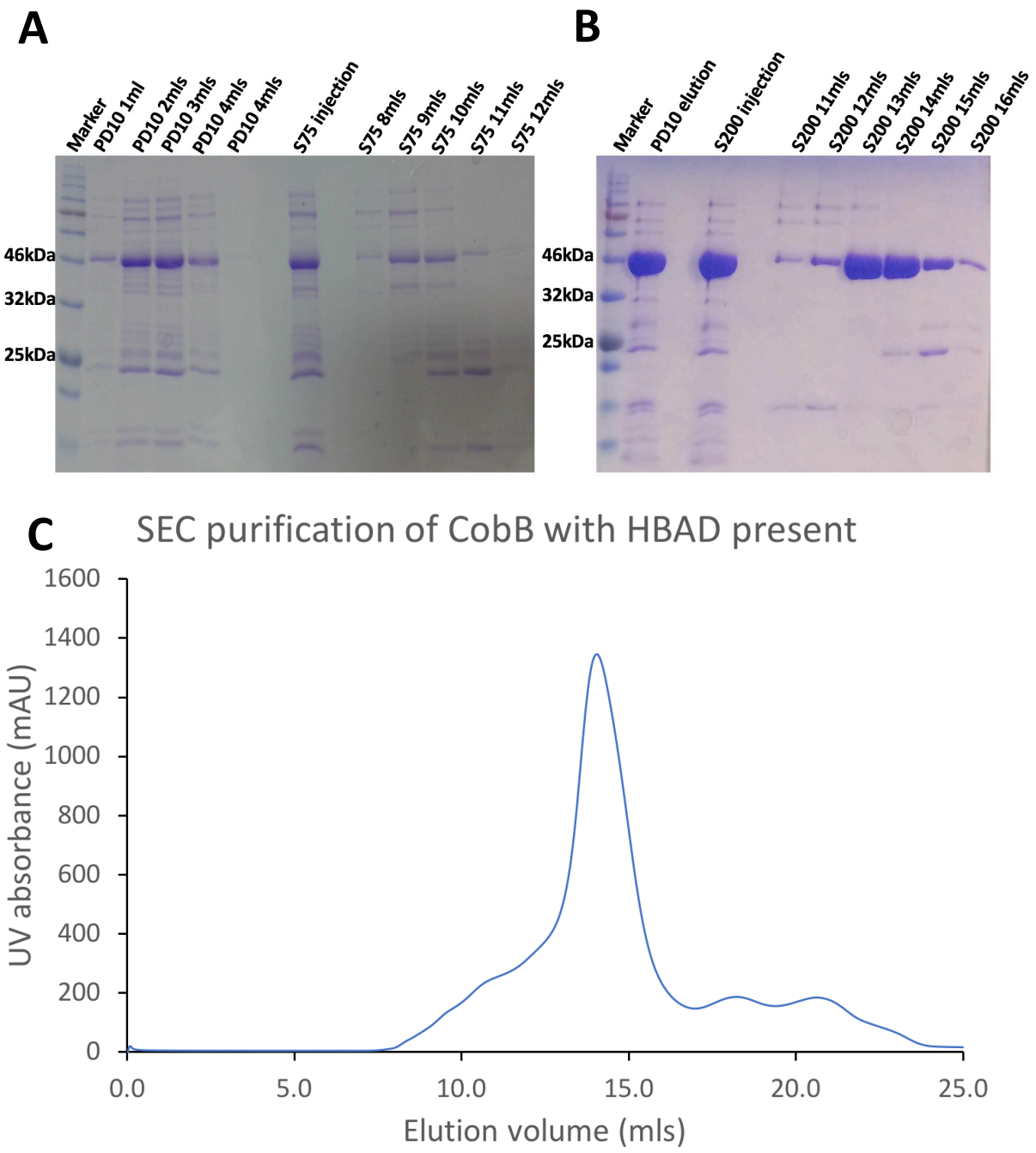


Figure 33: Purification of CobB (~45kDa) in presence and absence of ligand HBAD. A) SDS-PAGE analysis of the protein purification process with no ligands present. B) SDS-PAGE analysis of CobB with HBAD present C) UV profile of the S200 size exclusion column elution where the peak is from the fractions 11 to 16mls.

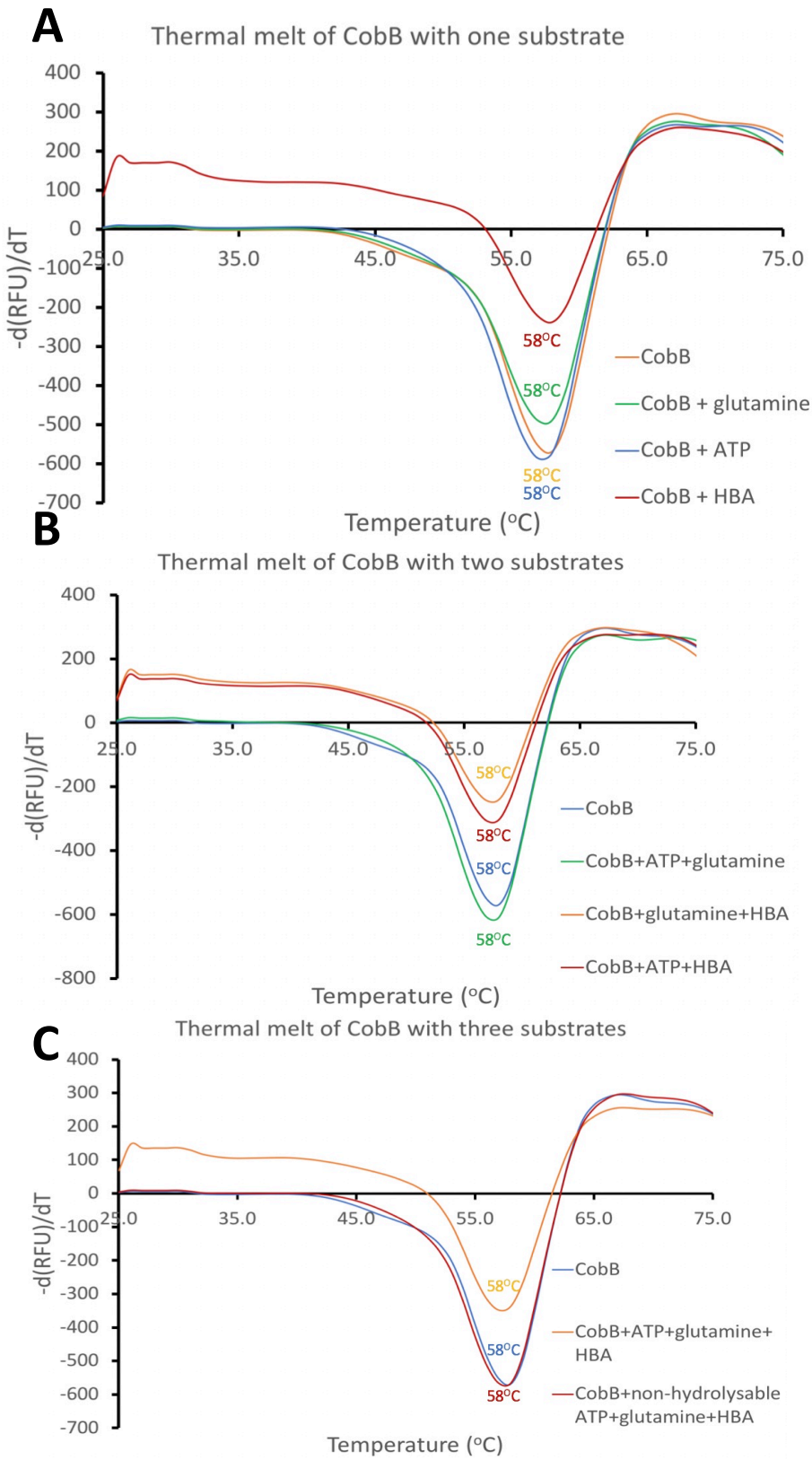


Figure 34: Real time PCR thermal melt studies of CobB in the presence of (A) one substrate only (B) two substrates (C) three substrates. These substrates included HBA, glutamine, ATP and non-hydrolysable ATP. The melting temperature (T_m) has been labelled on each curve.

3.3.2 CobB crystal screening and optimisation

Pure CobB protein was concentrated to 20mg/ml and put down for crystal screening conditions. Initial screening gave spherulites in multiple conditions, and any promising looking conditions were optimised further. Early optimisations were unsuccessful but when a microseeding technique described by Till et al, 2013 was used, crystals were obtained[104]. The best looking crystals were in the condition JCSG+ G2, in which the optimisation conditions ranged between 25%-30% w/v poly(acrylic acid sodium salt) 5100, 0.02M magnesium chloride hexahydrate and 0.1M HEPES pH7.5. These crystals are shown in Figure 35.

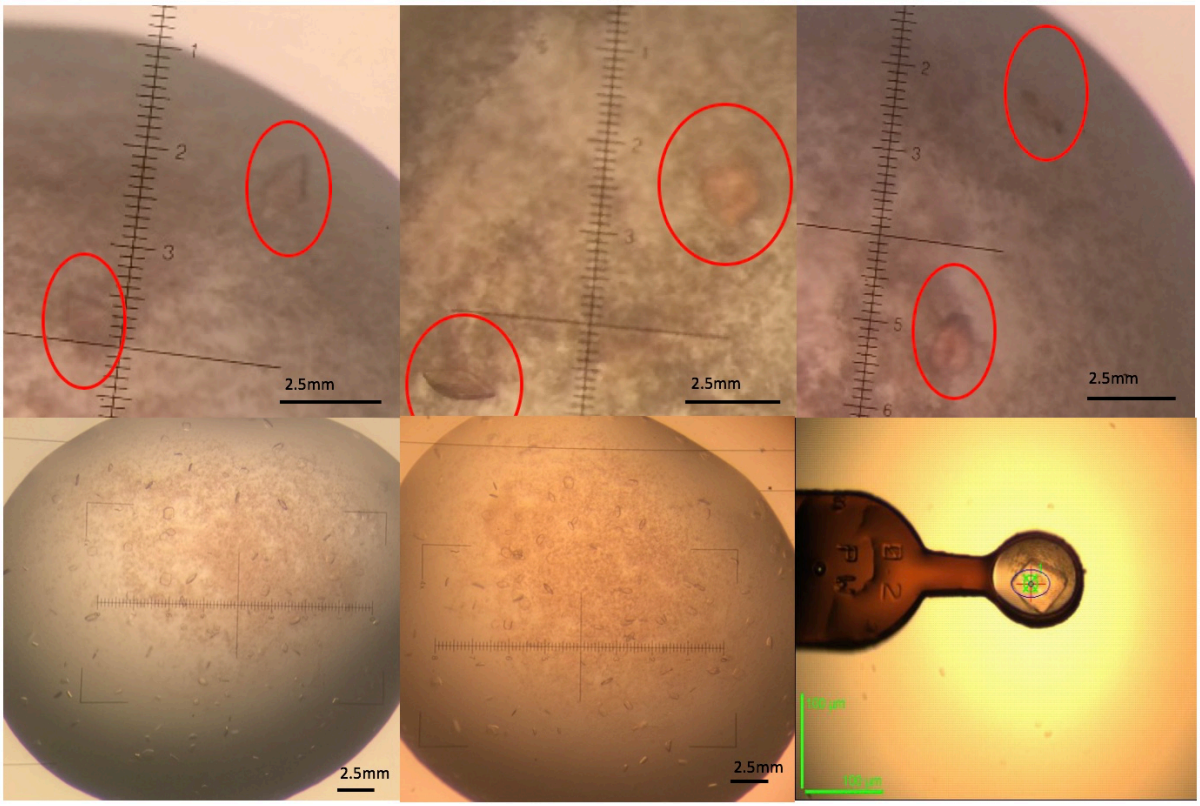


Figure 35: Crystals obtained of CobB from optimisation trays. The upper images show a large amount of precipitation, but red crystals have been highlighted. The lower images were from subsequent optimisation trays and despite adding HBAD the red colour was not observed and instead more crystals were produced, but they were smaller in size. The final image shows a loaded CobB crystal at the Diamond Synchrotron.

3.3.3 CobB crystal data processing

Diffraction data were collected at the Diamond Synchrotron to 3.0Å resolution from multiple trips and about 70 crystals, with collection strategies provided from the initial diffraction screens. From the processed data files, the best collection had a resolution of 3.0Å. These images were integrated using DIALS and scaled through AIMLESS in the CCP4 suite, as shown in Table 7. The space group selected for this reduction was P2 with a space group confidence score and Laue group confidence score of both 1.0.

Molecular replacement was unable to be used to solve the structure of CobB. Molecular replacement was carried out using the Phenix Suite with models which had been generated via the Phyre2 webserver. Despite generating 120 models all with a confidence score $\geq$98.7% the phases were unable to be solved and so a structure of CobB could not be generated.

CobB data collection statistics

Space group	P1 2 1		
Cell edges (Å)	55.090	113.725	67.194
	90.000	103.183	90.000
Wavelength (Å)	0.97625		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	3.00-113.72	9.00-113.72	3.0-3.18
Number of observations	107,199	4,130	17,608
Number of unique reflections	16,241	633	2630
Multiplicity	6.6	6.5	6.7
Rmerge	0.327	0.228	0.845
Rpim	0.138	0.093	0.358
Rmeas	0.335	0.247	0.92
Mean I/σ	4.9	6.4	3.3
$CC_{1/2}$	0.936	0.96	0.763
Completeness (%)	100	99.9	100
Wilson B estimate ($Å^2$)	25.72		

Table 7: Data collection statistics for CobB where Rmerge = Σ hkl Σ i |Ii–< I>| /– Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge R-factor = Σ hkl |Fo–Fc |/Σ hkl Fo, where Fo and Fc represent the observed and calculated structure factors, respectively[62,103].

3.3.4 CobB selenomethionine crystals

As it was not possible to solve the CobB structure by molecular replacement using diffraction data from four separate Synchrotron visits, an attempt was made to try and solve the phases using anomalous dispersion from incorporated selenium atoms. To do this CobB was produced and purified in minimal media where the amino acid methionine was substituted with selenomethionine (SeMet). HBA and ATP were added to the protein elution steps after the nickel column and size exclusion chromatography, to help stabilise the protein, since purifying CobB from HBAD producing cells had been shown to increase protein yields. As the protein was produced in cells grown in minimal media this did result in the expected lower yields and purity which is likely due to stress additional stress proteins synthesised in *E.coli*, Figure 36. However enough protein was purified and concentrated for setting up crystal trays with a concentration of 20mg/ml. The purification and crystal trays were set up in an identical fashion to a regular CobB purification in section 3.3.2 and produced small square plate crystals as shown in Figure 37. These crystals were extremely small and difficult to pick with a size no greater than 0.5mm. They were not red suggesting that HBAD was not present unlike the previous CobB crystals observed in section 3.3.2.

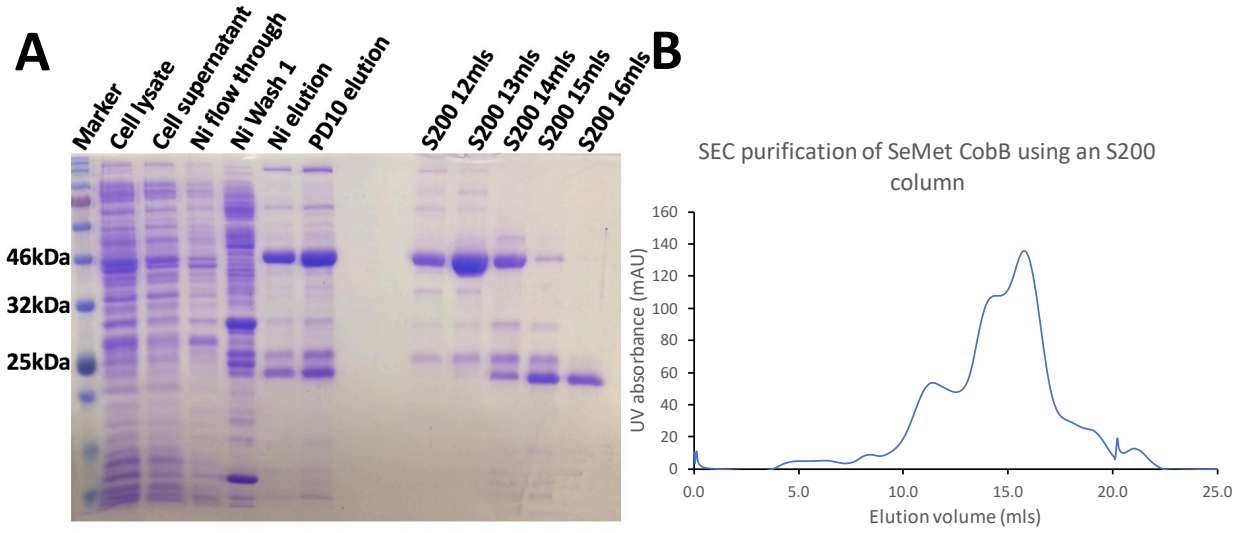


Figure 36: Purification of CobB labelled with selenomethionine (~45kDa). A) SDS-PAGE analysis of the protein purification process. B) UV profile of the S200 size exclusion column elution where the peak is from the fractions 12 to 16mls.

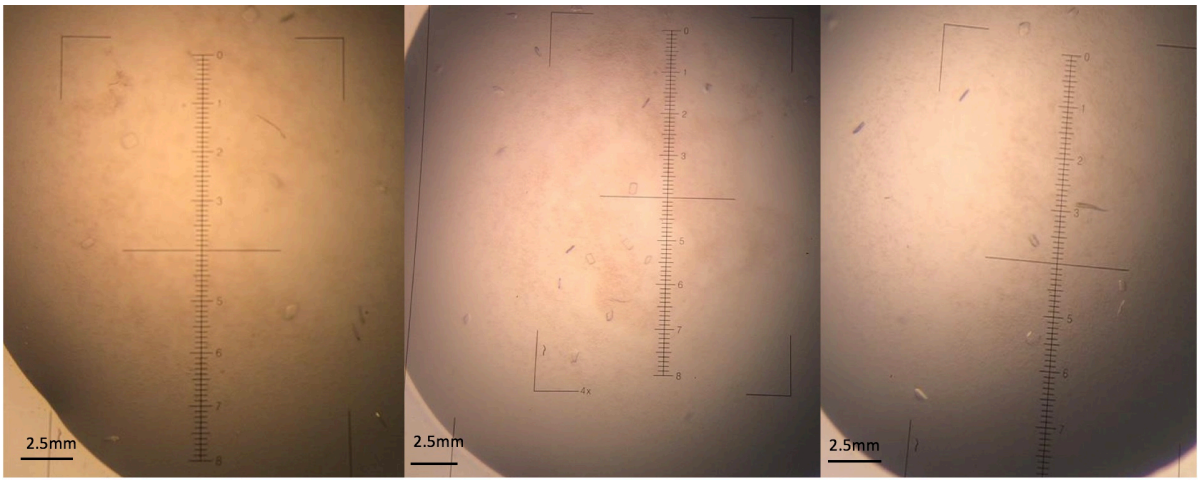


Figure 37: A selection of CobB SeMet crystal optimisations showing small flat square crystals.

During the diffraction data collection, the presence of selenium in the protein crystals was clear from the CHOOCH fluorescence scan that was carried out during the crystal screen process, Figure 38. The best diffraction data collected were from 1440 images generated from 720^{o} rotation with 0.5^{o} oscillations and gave a resolution up to 3.03Å. The space group and unit cell dimensions were identical to those obtained for non-labelled CobB crystals and the data reduction statistics are summarised in Table 8.

In an attempt to solve the phases via the selenium incorporation Parrot via CCP4-2i and AutoSol via Phenix were both trialled. The electron density maps generated were not of sufficient quality to attempt to trace the protein backbone. This was highlighted by a mid-slope of anomalous probability of 0.882, which was less than 1 meaning the location of selenium atoms in the crystal diffraction data could not be accurately placed.

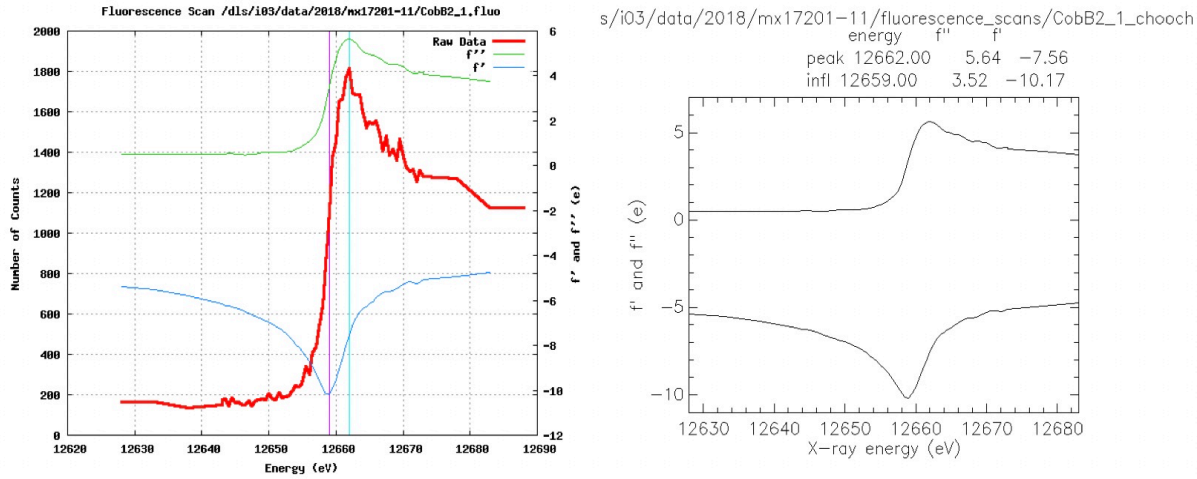


Figure 38: CHOOCH fluorescence scan showing the presence of selenium in CobB crystals labelled with selenomethionine.

SeMet CobB data collection statistics

Space group	P 1 2 1		
Cell edges (Å)	55.947	113.734	68.076
	90.000	103.365	90.000
Wavelength (Å)	0.97925		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	3.03-113.73	9.09-113.73	3.03-3.21
Number of observations	212,673	7,743	34,370
Number of unique reflections	16,226	629	2,626
Multiplicity	13.1	12.3	13.1
Rmerge	0.577	0.107	2.262
Rpim	0.165	0.029	0.647
Rmeas	0.601	0.041	0.647
Mean I/σ	5.2	14.2	2.0
$CC_{1/2}$	0.981	0.995	0.531
Completeness (%)	100	99.9	100
Anomalous multiplicity	6.5	6.4	6.5
Anomalous completeness (%)	99.9	99.6	100
DelAnom $CC_{1/2}$	0.01	0.048	0.038
Mid-slope of Anom Probability	0.882		
Wilson B estimate ($Å^2$)	44.73		

Table 8: Data collection statistics for CobB labelled with selenomethionine where Rmerge = Σ hkl Σ i |Ii–< I>| /– Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge R-factor = Σ hkl |Fo–Fc |/Σ hkl Fo, where Fo and Fc represent the observed and calculated structure factors, respectively [62,103].

3.3.5 CobB co-crystallisation with CobH

As CobH exhibits product trapping activity, it was envisaged that CobH and CobB could be co-crystallised together showing a trapped interaction which initiates the release of HBA. To attempt this CobH and CobB were purified and put down together for crystal screens in the presence of 10x excess non-hydrolysable ATP, adenosine 5'-(β,γ-imido)triphosphate lithium salt hydrate, to prevent HBA turnover. A typical CobH protein purification is shown in Figure 39. The crystal screening resulted in multiple conditions yielding red/orange square like crystals, as shown in Figure 40. These crystals were numerous and large enough to be picked directly from the screen droplets, without the need for optimisation. Cryo-solutions were made by using the existing well conditions and diluting with 20% glycerol if a cryogenic agent was not already present, before being sent off to Diamond Synchrotron for diffraction data collection.

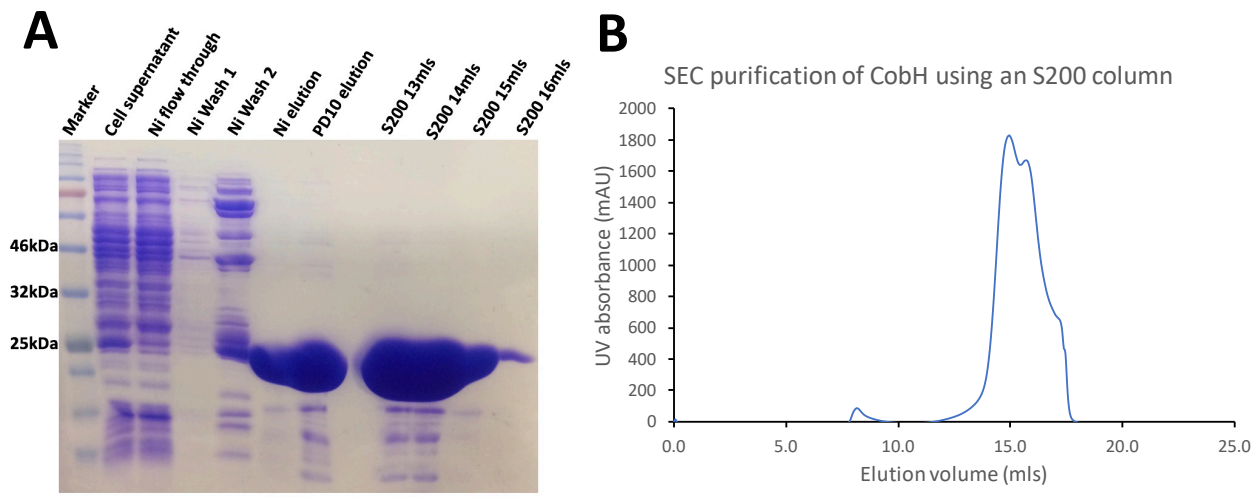
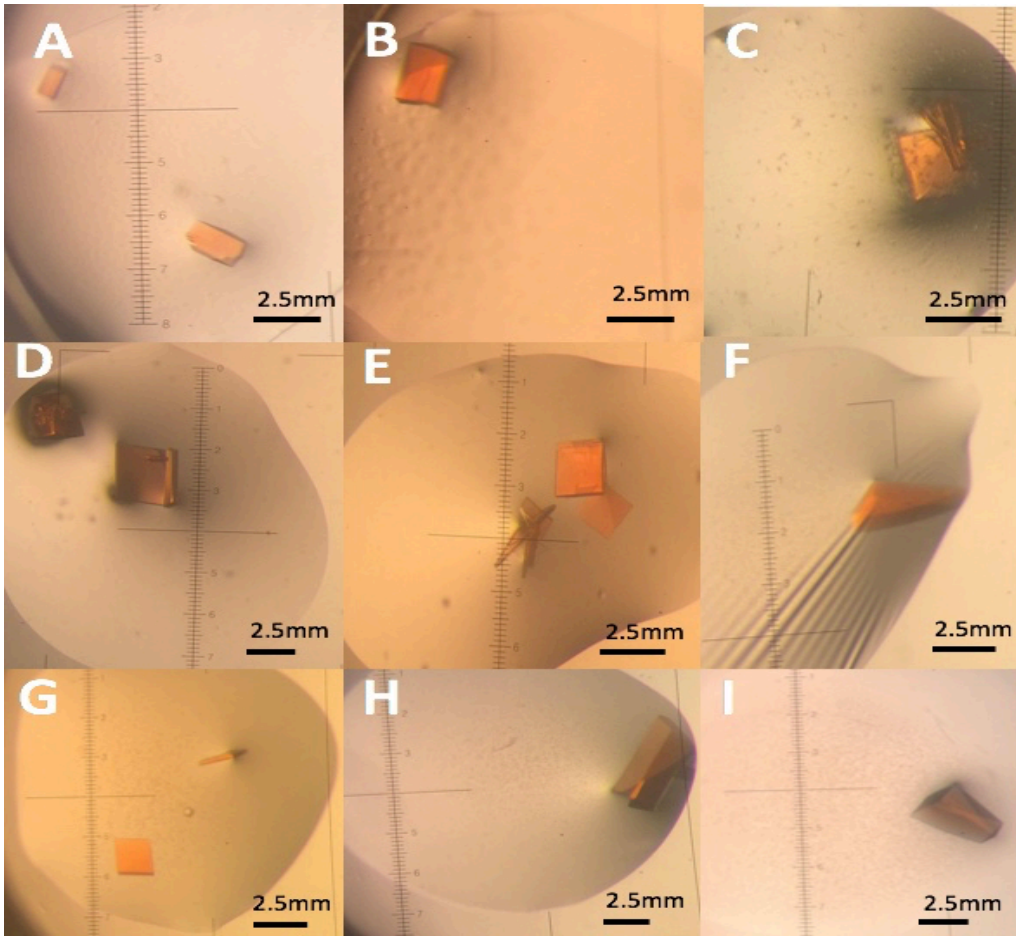


Figure 39: Purification of CobH (~25kDa). A) SDS-PAGE analysis of the protein purification process. B) UV profile of the S200 size exclusion column elution where the peak is from the fractions 13 to 16mls.



Panel	Crystal screen	Crystal Conditions
A	JCSG+ 2-12:	0.1M imidazole pH8.0, 10% w/v PEG 8000
B	JCSG+ 2-31	0.1M succinic acid, 20% w/v PEG 3350
C	PACT E1	0.2M sodium fluoride, 20% w/v PEG 3350
D	PACT E4	0.2M potassium thiocyanate, 20% w/v PEG 3350
E	PACT E5	0.2M sodium nitrate, 20% w/v PEG 3350
F	MemGold 2 2-6	0.2M calcium acetate, 0.1M sodium acetate pH5.0, 38% PEG 400
G	MemGold 2 2-15	0.3M barium chloride, 0.1M MES pH6.5, 34% v/v PEG 400
H	MemGold 2 2-16	0.32 lithium chloride, 0.1M sodium citrate pH5.5, 14% w/v PEG 4000
I	SSII-40	2.0M sodium chloride, 10% w/v PEG 6000

Figure 40: A selection of crystals observed when CobB and CobH were put down together from a range of crystallisation screen conditions, with a neighbouring table summarising the conditions.

During the diffraction data collection, the automated collection strategies showed a smaller unit cell and different space group to previously described CobB data collections. Molecular replacement confirmed that the only protein present in the crystals was CobH. The best collection was solved to 1.5Å and the crystal data statistics are summarised in Table 9. The diffraction images were scaled using Dials via CCP4-2i suite and molecular replacement carried out using Phaser in the Phenix suite, with CobH (PDB 4FDV) as a model.

CobH data collection statistics

Space group	C 1 2 1		
Cell edges (Å)	70.062	65.622	48.453
	90.000	99.297	90.000
Wavelength (Å)	0.91587		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	1.50-47.82	8.22-47.82	1.50-1.53
Number of observations	97,975	735	3,901
Number of unique reflections	34,370	224	1,663
Multiplicity	2.9	3.3	2.3
Rmerge	0.056	0.045	0.73
Rpim	0.069	0.028	0.574
Rmeas	0.069	0.053	0.932
Mean I/σ	8.5	23.5	1.2
$CC_{1/2}$	0.997	0.998	0.52
Completeness (%)	99.2	98.6	97.7
Wilson B estimate ($Å^2$)	18.77		

Refinement statistics

Resolution (Å)	1.50-47.82
R-work/R-free	0.19/0.21
Number of atoms (protein/ligand/water)	1/1/212
Rmsd bond length	0.08
Rmsd bond angle	1.514
Mol Probability	0.96
Residues in most favoured regions (%)	98.54
Outliers (%)	0

Table 9: Crystal data collection and refinement statistics for CobH & CobB co-crystallisation, which ended up being just CobH. Rmerge = Σ hkl Σ i |Ii–< I>| /– Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge R-factor = Σ hkl |Fo–Fc |/Σ hkl Fo, where Fo and Fc represent the observed and calculated structure factors, respectively. The R-Factor is calculated using 95% of the data included in refinement and R-free the 5% excluded[62,103].

The structure of CobH was refined with HBA bound and solved to 1.5 Å which is a small improvement in the resolution on five of the six existing structures available in the Protein Data Bank, in which only one has been solved to a higher resolution of 1.45Å[36,53,67]. The degree to which this structure of CobH compares to the existing HBA product bound structure 4FDV.pdb is identical as highlighted by overlaying the two structures in Figure 41, with a root-mean-square-deviation (RMSD) of 0.071 when the alpha carbons of the protein backbone where aligned.

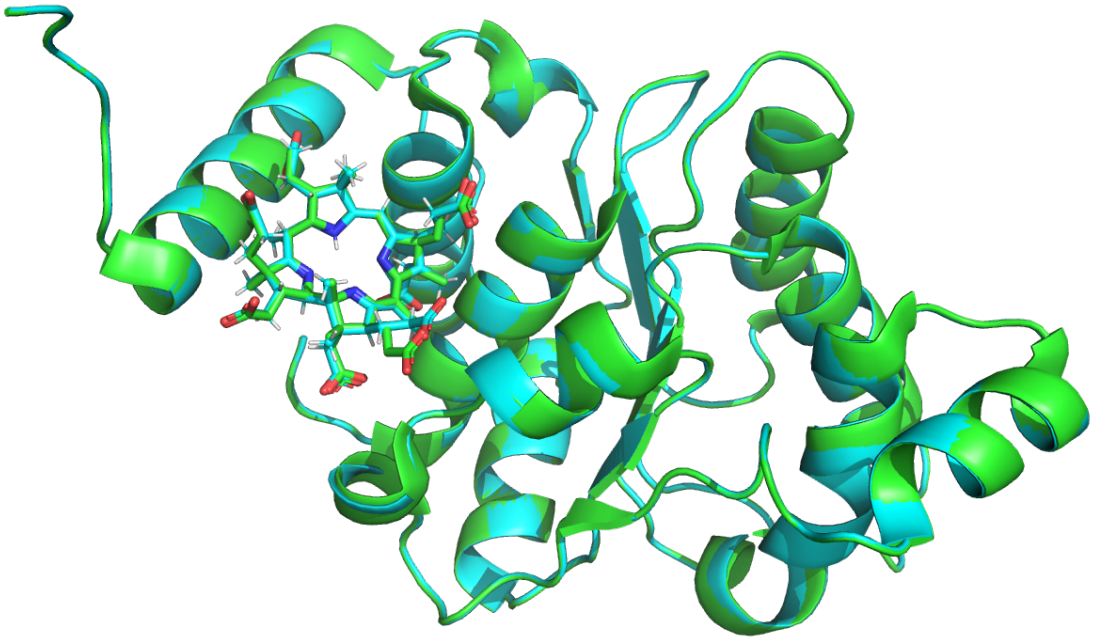


Figure 41: The CobH structure obtained with HBA bound (blue) overlaid with the existing CobH with HBA structure 4FDV.pdb (green), showing little difference between the two. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC. When 209 residues were aligned on Pymol using the function 'align Prot1////CA, Prot2, object=alignment' a RMSD of 0.071 was generated from 200 of these residues.

3.4 Discussion and future work

3.4.1 CobL crystallography

During the course of this project it has not been possible to solve the structure of CobL from the present series of experiments undertaken. Crystals were obtained during screening and optimisations were carried out. However, despite multiple crystallisation conditions trialled, they continued to result in a large number of very small low-quality crystals. These crystals proved difficult to physically pick and troublesome to centre on the synchrotron beamline.

A number of suitable models for molecular replacement of CobL were available and these included the C-terminal domain of CobL (PDB 3NJR)[53] and the anaerobic equivalent enzyme CbiT (PDB 1KXZ)[65]. Given the high degree of similarity of the C-terminus, therefore it could be concluded that the full length structure of CobL is structurally very similar to the anaerobic enzymes CbiET, of which the laboratory group currently also has an unpublished structure. Alternatively, given the similarity of the N-terminus of CbiET to CbiE it could be concluded that other canonical methyltransferases which have already had their structures solved would also be appropriate models as they are all Class III methyltransferases.

Upon the reduction of CobL diffraction data the multiplicity and completeness were good. However, the limitations of the data collection arose from the high resolution of the outer shell of 3.6 Å. Despite an overall Rpim score of 0.087 this is misleading as the outer shell had a very high Rpim of 0.251, which despite containing a correction for redundancy is still similar to Rmerge indicating a poor signal-to-noise ratio[105]. Finally, the L-test indicated that the CobL crystals were highly twinned, with an almost perfect twin fraction score of 0.45. This meant that it was not possible to solve the full-length structure of CobL using molecular replacement, despite having many promising models readily available.

During the crystallisation process of CobL it was observed that more crystals were produced when CobL was in the presence of the ligand SAH, which supported the thermal melt studies of CobL stability with various ligands. Based on the crystal structures of CobL^{c} and CbiET, these observations

agree with the binding and docking interactions present in these structures. SAH is present in both of these structures and is buried deep in the active site, Figure 42. The unpublished structure of CbiET also contains the anaerobic equivalent tetrapyrrole cobalt-PC6B, which does not appear to bind as tightly, which is in further agreement with the thermal melt studies where PC6B had no effect on the melting temperature (T_m) and therefore protein stability of CobL.

Clearly more time needs to be spent on producing larger and better-quality crystals. This could be achieved by further screening trials and continued optimisation. Attempting to produce crystals of CobL protein from a different species of bacteria may result in better crystals, as existing efforts have already been pursued. Given that small crystals can be produced it is possible that the XP Screen: Protein Crystal Glue produced by MiTeGen may assist in crystal production and quality, as summarised in Figure 43. This screen contains Anderson-Evans polyoxotungstate $[TeW_6O_{24}]^{6-}$, which also offers the option to carry out anomalous dispersion diffraction data collections for phasing due to the presence of six tungsten atoms.

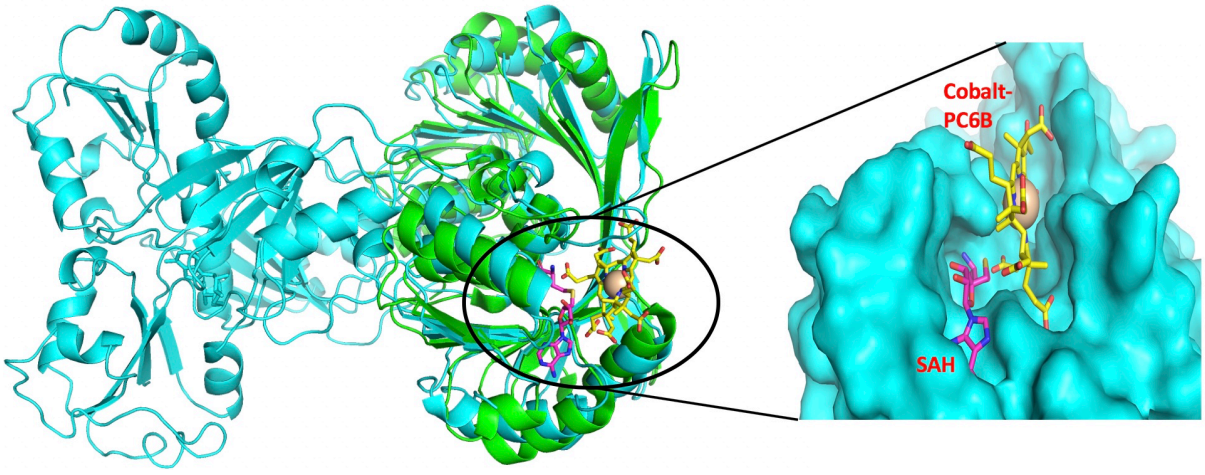


Figure 42: The protein structure of CobL^c dimer in green (3NJR.pdb) overlaid the CbiET dimer in blue (unpublished) to show the structural homology. SAH (purple) can be seen buried in the active site, whereas cobalt-PC6B (yellow) is more exposed when bound at the C-terminus when a surface view is applied to CbiET and rotated 90^0. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

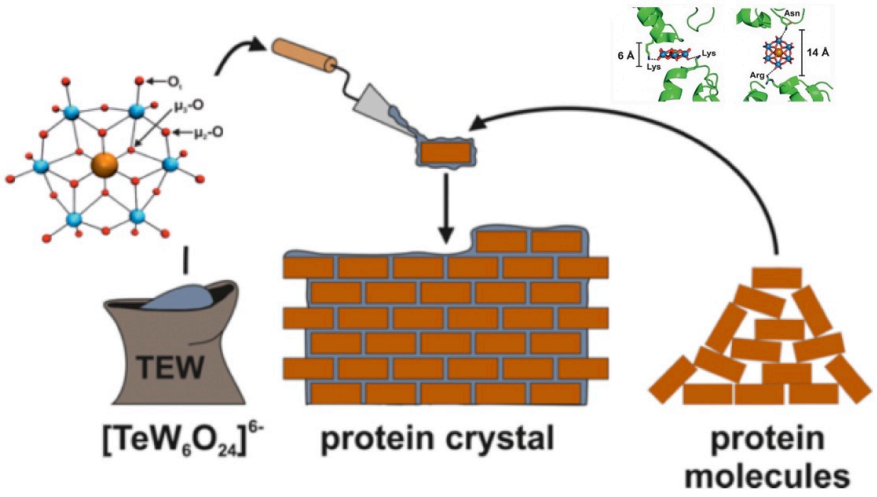


Figure 43: A simplified explaination of how XP Screen: The Protein Crystal Glue-96 solutions MiTeGen works from their website.

3.4.2 CobK crystallography (CobL co-crystallisation with CobK)

A higher resolution structure of the native form of CobK was obtained in an attempt to co-crystallise a CobK-CobL complex. Three new features were observed during the refinement process. This included the lack of the disulphide bond presented in the known published structure. Given CobK is a bacterial cytoplasmic enzyme it is unlikely a disulphide bond would be able to form in a reducing environment. A metal ion was observed in the structure via the octahedral geometry between neighbouring amino acids and nearby water molecules. However, it was concluded that it was likely to be a sodium atom and was observed as an artefact of the crystallisation conditions which contained 1M sodium acetate. As this occurred in the NADPH binding site it could also explain why no substrate binding was observed when excess ligands were added to the screening process. Finally, this structure was in a different space group of $P2_1$ compared to the current PDB structures which are in $P2_1\,2_1\,2_1$. The crystal packing present was much denser with two protein chains orientated back to back and inverted with respect to one another in the asymmetric unit. This could have potentially come about due to a transisent interaction between CobK and CobL resulting in different contacts present during crystal packing and growth. With no ligands present there was an

improvement in the resolution reported of the already existing apoenzyme. It was interesting that the flexible loop region of CobK could be observed in one of the protein chains due to crystal contacts trapping it and allowing this dynamic loop to be modelled. When this protein chain was overlaid with the holoenzyme available in the Protein Data Bank, there was little difference between these structures as shown in Figure 44. When this structure is overlaid with existing CobK structures it bears a closer resemblance to the apoenzyme (5C4R.pdb) with an RMSD of 0.534 and includes the active site loop residues. The degree of variation in the flexible loop region, compared to the product structure, and lack of electron density in the apoenzyme highlights just how flexible this region is. For a structure solved to 2.0Å an R-free value of 0.31 is higher than expected given an R-work value of 0.24. Clearly the apoenzyme is more flexible in nature than the product-bound form, which may have resulted in more variation within the crystal packing and flexible loop regions. Despite this the model is in good agreement with existing structures.

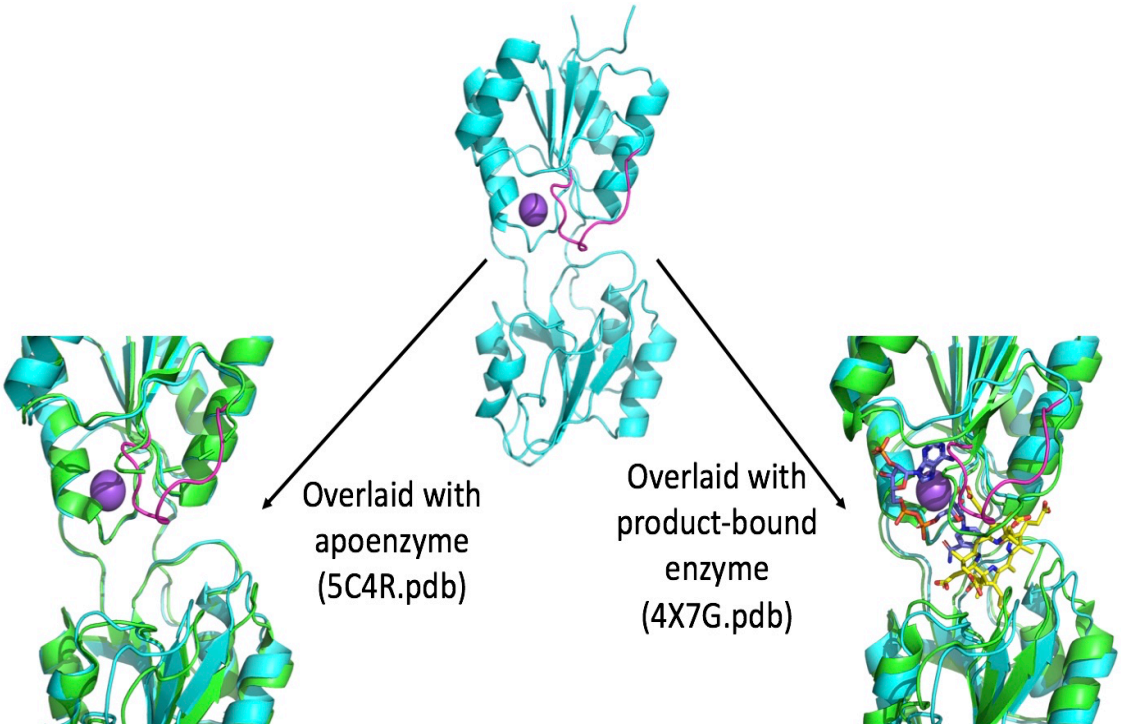


Figure 44: The CobK in blue is the structure solved within this project, where the flexible loop region and sodium ion have been coloured purple. The blue CobK has been overlaid with the existing apoenzyme (green) with an RMSD of 0.534 between 211 alpha carbon atoms. The product bound CobK (green with PC6B in yellow and NADPH in blue) gave an RMSD of 1.228 between 235 alpha carbons. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.The Pymol alignment function was 'align Prot1////CA, Prot2, object=alignment'.

3.4.3 CobB crystallography

During the purification of CobB in the presence of HBAD, product binding was visibly observed by the pink pigmentation co-migrating during nickel column chromatography, size exclusion chromatography and concentration steps. This also happens with CobK and CobH protein purifications with their respective products. HBAD binding to CobB appeared to show an increase in protein stability and yield but when this was measured using a thermal melt study surprisingly there was no change in the melting temperature of CobB.

In a similar way to the CobL crystallography, the structure of CobB remains unsolved despite crystals being obtained. The highest resolution data obtained was limited to only 3.0Å, but given models were generated with extremely high confidence scores, it still was not possible to solve the structure of CobB using molecular replacement. As was the case with CobL, a high Rpim of 0.138 was obtained despite a good multiplicity and completeness score, but once again the outer shell Rpim was similar to Rmerge highlighting poor signal-to-noise ratio[105]. The quality of the crystals was insufficient as indicated by the $CC_{1/2}$ of 0.763 in the outer shell. The models generated for molecular replacement were from a varied range of existing PDB structures available, combined with extremely high confidence scores, which were in line with the ATP and glutamine binding domains present in class I glutamine-dependent amidotransferases. The Matthews coefficient estimated that two protein chains would likely occupy the unit cell, which is in agreement with the literature of CobB forming a dimer[51]. There was also the presence of non-crystallographic symmetry in the asymmetric unit cell which may also have hindered molecular replacement strategies.

Despite producing additional crystals containing selenium labelled methionine, which were successfully picked for data collection, the structure of CobB could still not be solved. The selenium anomalous crystal data was again restricted to a high resolution limit of 3.03Å. These data had an anomalous completeness of 99.9% and anomalous multiplicity of 6.5 but once again a high Rpim of 0.165 and $CC_{1/2}$ of 0.531 highlight the low quality of diffraction for these crystals, as well as a low mid-slope anomalous probability score of 0.882 which meant the selenium atoms could not be

assigned correctly. It is possible that prolonged radiation damage decreased the quality of the data over this time course as diffraction data had to be collected to 720°. Also given the nature of the small flat square crystals it was difficult to keep the beam line centred down the planar edge, which will have contributed to a lower quality diffraction data collection at various oscillations.

It was surprising to discover that there was a structure of cobyrinic acid a,c-diamide synthase (anaerobic CobB equivalent) already available on the PDB (4PSF, 5IF9 and 5IHP). Upon further investigation this protein structure appears to only contain the ATP binding domain. The Phyre2 modelling server did produce a model based on this PDB structure but it was ranked at 54/120 with 24% sequence similarity and had a confidence score of 99.5%. Despite these available PDB structures and the Phyre2 models, molecular replacement phasing was still unable to solve the structure of CobB. This structure of cobyrinic acid-a,c-diamide originated from *Mycobacterium smegmatis*. Upon locating this gene from its genome which has locus tag 'MSMEG_1927', the neighbouring gene sequences did not show the presence of glutamine binding domain with 'MSMEG_1926' a phosphoglycerate mutase (glycolysis protein) and 'MSMEG_1928' a serine/threonine phosphatase. This could mean that the glutamine binding domain is located elsewhere in the genome of *M. smegmatis* and, in a similar fashion to CbiET, indicate CobB arose as a result of a gene fusion event. Alternatively, the amidation steps of cobyrinic acid may occur via the neighbouring phosphatase gene, in which this enzyme subsequently removes the phosphate group added via the ATP binding domain, highlighting an alternative catalytic mechanism.

If a structure of CobB had been obtained further studies would mimic those in line with other class I glutamine-dependent amidotransferases [76,106–109]. These studies typically go on to identify the amino acid residues which are involved in the formation of the molecular tunnel required for the safe delivery of ammonia from the glutamine binding site to the ATP binding site[107,108]. Additional studies have also attempted to explore how these two binding sites are linked and coordinate their respective reactions when both are occupied by substrates to avoid the loss of wasteful hydrolysis[76,109]. The antibiotic acivicin has also been shown to be an inhibitor in glutamine-

dependent proteins, as it is able to bind to the active site cysteine and has assisted in crystallography structure determinations [110,111]. But most importantly, it would have been possible to study the CobH-CobB pair structurally to better understand product/substrate transfer.

As discussed above with CobL crystallography and future work in section 3.4.1 continued crystallisation screening and optimisation may produce higher quality and larger crystals and result in better diffraction data. Again, it could be worth trialling the crystal glue provided by MiTeGen or the antibiotic acivicin. Alternatively, the use of mercury soaks, given CobB contains a readily available active site cysteine, would produce a much stronger anomalous signal when collecting crystal diffraction data. Finally using a CobB protein from another species may be easier to purify and crystallise, yielding a near atomic resolution structure.

3.4.4 CobH crystallography (CobB co-crystallisation with CobH)

There is little new information to report on the 1.5Å structure of CobH solved during this project. There was a very small improvement on the resolution but otherwise this structure is identical to existing product bound structures. CobH was a relatively easy protein to purify in high amounts and produced numerous large pink crystals in a variety of crystallography screen conditions in the presence of its product HBA.

The co-crystallisation attempts did not assist in the crystallography of a CobB structure. A preliminary thermal melt study of CobH and CobB was carried out to probe if there may be an interaction between these enzymes as shown in Figure 45. The presence of HBA did interfere with the absorbance measurements of SYPRO-orange used during this technique, meaning a change in melting temperature (T_m) of CobH could not be measured. When a thermal melt was carried out on CobH and CobB together, with CobB activity being controlled by using either ATP or non-hydrolysable ATP, the presence of HBA/HBAD signal continued to interfere. However, what was interesting to observe is that when PC6B (which does not absorb in the absorbance range of SYPRO-orange) was converted into HBA and subsequently HBAD in the presence of CobL, CobH and CobB together, the HBA signal was lost, Figure 45C. The HBA signal is produced when SAM is added for

CobL to convert PC6B into PC8 which can be subsequently converted in HBA by CobH and non-hydrolysable ATP prevents CobB activity. Yet when this reaction is set up and repeated with ATP for CobB to convert HBA into HBAD, the HBA signal is not observed. This may indicate that a protein interaction is occurring and blocking the HBA absorbance measurements at lower temperatures. The results of Figure 45B could possibly be explained by the fact that thermal melt studies are carried out in an excess of substrates, so unbound HBA interfered, whereas in Figure 45C HBA can only be produced directly from PC6B and therefore proportional to enzyme activity. There are a lot of components in some of these thermal melts and these experiments are very preliminary, so it is certainly not a conclusive result that these proteins interact, but it is an interesting observation regardless.

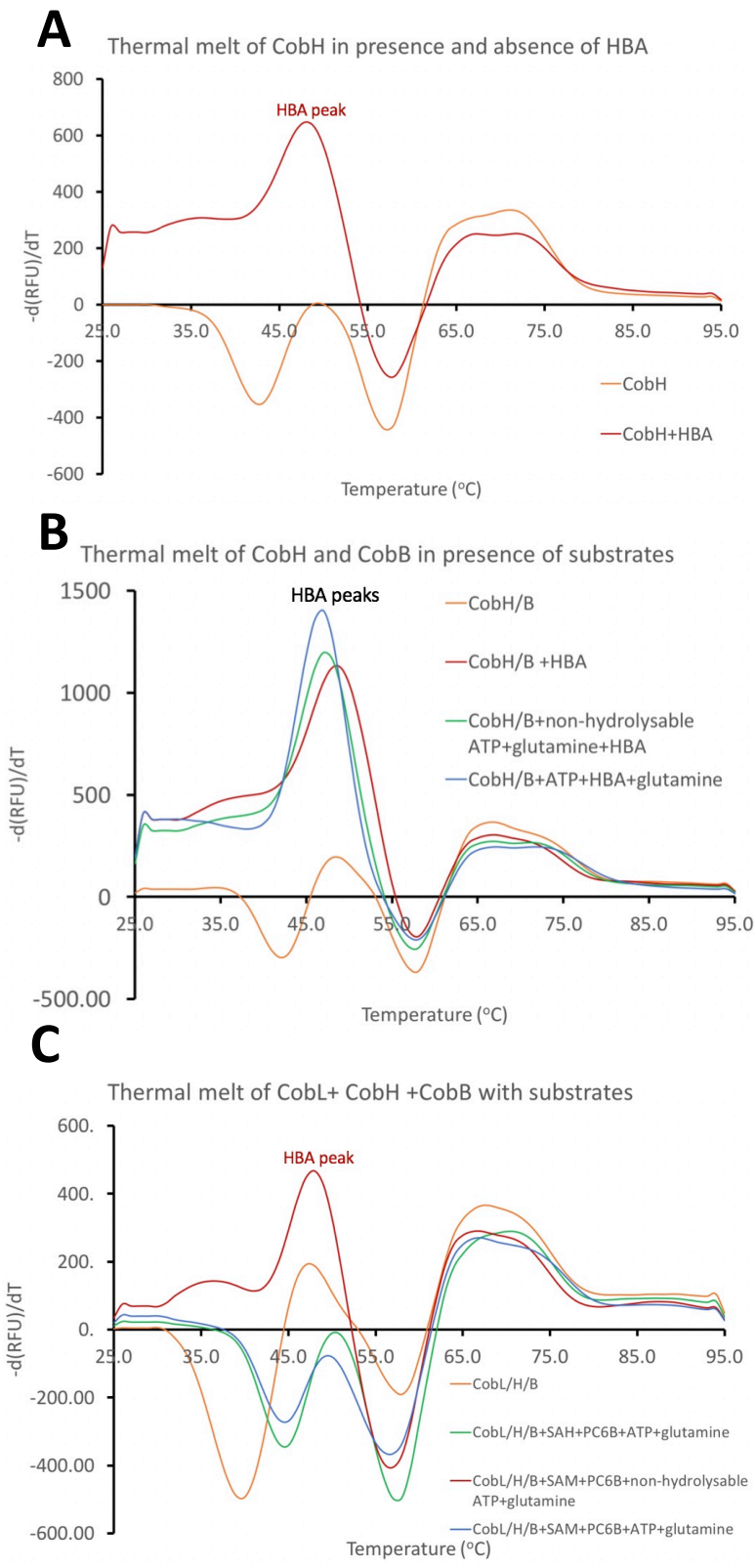


Figure 45: Real time PCR thermal melt studies of (A) CobH in the presence and absence of HBA (B) CobH and CobB in the presence and absence of HBA and where CobB activity is controlled using ATP/non-hydrolysable ATP (C) CobL, CobH and CobB where PC6B is converted into HBA using SAM and HBAD using glutamine and ATP.

3.4.5 Concluding remarks

Despite numerous efforts to produce and optimise CobL and CobB crystals, it was perplexing that the crystals produced and sent off for data collection were of insufficient quality for molecular replacement or SeMet anomalous dispersion experiments. Co-crystallisation efforts of the enzyme pairs were promising but unfortunately resulted in the crystallisation of only one of the present proteins, in which both of these had already had their structures solved and published. The structures of CobK or CobH that were solved did not reveal anything substantially new regarding how these enzymes function and only really offered a small improvement in the resolution.

There are certainly plenty of continuing efforts that can be undertaken to produce good quality crystals to help solve the structures of CobL and CobB but due to the time constraints of this project they could not be acted on. In an effort to explore the enzyme pairs and how the subsequent enzyme in the cobalamin pathway receives its substrate it was considered that perhaps other techniques to probe the likelihood of a protein-protein interaction could be used as discussed in the next chapter.

Chapter 4: Investigating product binding and release from CobK

4.1 Introduction, aims and objectives

The crystal structure of CobK with NADPH and its product precorrin-6B (PC6B) was solved by the Pickersgill group in 2015[62], which in fact surprisingly contained PC6B which was additionally methylated at C5. As described in section 1.4.2 the structure shows tight binding of the product PC6B, via the closure of a flexible loop region over the active site. It is plausible to suggest that product binding during the cobalamin biosynthetic pathway prevents off-pathway reactions. To explore and confirm product binding activity in CobK the following experiments were carried out:

1. UV-VIS spectroscopy to follow the purification of PC6B and how the absorbance changes under different buffer conditions and during the addition of CobK via a titration experiment.
2. Thermal melt studies of CobK and how its stability changes in the presence of the ligands NADPH and PC6B by calculating the change in melting temperature (T_m).
3. 2D nuclear magnetic resonance (NMR) using N^{15} labelled CobK protein to observe how the spectra change in the presence of the ligands NADPH and PC6B.
4. Isothermal calorimetry (ITC) to quantify the dissociation constant (K_d) for PC6B product binding.

After demonstrating that CobK binds its product tightly, the next question tackled is how the cobalamin pathway continues and CobL, the subsequent enzyme in the pathway, obtains its substrate PC6B. There are two possibilities; either the dissociation constant (K_d) is sufficient for the

CobL catalysed methylations to proceed at sufficient rate, or there is some active transfer process between CobK and CobL. The latter prevents the possibility of off pathway reactions and wasteful loss of intermediates half way through this biosynthetic sequence. An exciting possibility is that a protein-protein interaction occurs between CobK and CobL to facilitate the successful delivery of PC6B between the two enzymes. X-ray crystallography of the two proteins was attempted and discussed in the previous chapter, but this did not lead to a successful co-crystallisation of a CobK-CobL complex. Therefore, other laboratory-based techniques were used to explore release of PC6B to CobL including:

1. Native-PAGE gels, a nickel column pull down and size exclusion chromatography (SEC).
2. UV-VIS spectroscopy titration to explore if SAM or SAH triggers the release of product PC6B.
3. Kinetics of PC6B turnover in the presence and absence of CobK.
4. 1D proton NMR to observe if CobK and the CobL C-terminus ($CobL^c$) potentially interact.

Of the four enzymes CobK, CobL, CobH and CobB investigated in these last two chapters, two of these, CobK and CobH, have demonstrated visually, experimentally and structurally product binding activity[53,62,67]. Why these two enzymes, which are so close together in the cobalamin biosynthetic pathway, do this is unclear and it raises the question as to whether there is a purpose for this? Having already discussed the possibility of substrate channelling to prevent the wasteful loss of intermediates, in what is a complex and metabolically demanding biosynthetic pathway, it is impressive that specificity is maintained given the high degree of similarity between the various cobalamin intermediates. In this case substrate channelling may also serve a purpose of controlling the order of reactions and prevent out of turn reactions. Failure to methylate C20 via the enzyme CobI, alters the chemistry of biosynthesis in such a way that CobJ, only two steps further along, is then unable to correctly carry out the corrin-ring contraction, resulting in a dead-end intermediate[56].

For example it is reasonable to suggest that the cobalamin intermediate PC6B may be able to bind to CobB, subsequently missing out decarboxylation and methylation of C12 via CobL and CobH which could be why CobK ensures delivery directly to CobL as discussed above via a protein-protein interaction. In a similar fashion CobH may deliver HBA to CobB for the successful di-amidation reaction. CobQ is an enzyme further upstream in the pathway and also carries out amidation reactions in a very similar fashion to CobB, which could therefore interfere with the enzyme reactions between these two proteins.[45]. To explore incorrect/correct binding of cobalamin intermediates to CobK the following were carried out:

1. Fluorescence spectroscopy to demonstrate if hydrogenobyrinic acid (HBA), the product of CobH, can bind to CobK.
2. An attempt to crystallise CobK with PC6B, as the existing structure contains an over-methylated ligand of C5-methylated-PC6B.

4.2 PC6B binding and release to CobK

4.2.1 PC6B purification

The purification of PC6B fully relies on the product binding activity of CobK, as it is isolated from CobK after a nickel column purification step. For this reason, an additional cell pellet containing over expressed CobK was added to the PC6B producing cell pellets so that any additional free PC6B can be mopped up, given these cell cultures are grown for a 24-hour period. Once the cells have been sonicated and the supernatant collected, this is run down a nickel column where CobK with bound PC6B binds to the nickel beads via its histidine tag.

Upon collecting the nickel column elution (which was brown in colour) the pH was dropped from pH7.5 to pH4.0. This resulted in a very clear colour change from brown to yellow, accompanied by some precipitation which was removed by centrifugation Figure 46. The precipitation was shown to be CobK, confirmed via an SDS-PAGE gel and the colour change of brown to yellow was assumed to be PC6B changing from a protein bound state to a free state, which was confirmed by UV spectroscopy, shown in Figure 46. Interestingly when the pH was subsequently returned to pH7.5 a second occurrence of protein precipitation occurred, which again was confirmed to be CobK, via an SDS PAGE gel, and this was removed via centrifugation. This second wave of precipitation removed most of the remaining CobK protein present, which was important for improving PC6B purification. Changing the pH twice proved to be a significant addition to the purification as it ensured the removal of all of CobK, when compared to the original purification protocol provided, shown in Figure 47. For the reverse phase column step to work it was essential to drop the pH back to pH4.0, upon which no further precipitation was observed, to ensure that PC6B would bind to the column in its protonated form. The PC6B could clearly be seen binding the reverse phase column, due to its yellow colour, and a UV scan was used to confirm the presence of PC6B in the elution, Figure 48. Mass spectrometry was also used to confirm PC6B and the m/z agreed with that recorded in Deery, et al, 2012[53], Figure 49. The reverse phase column elution was freeze dried via lyophilization. The PC6B powder was weighed out and stock solutions prepared for future use.

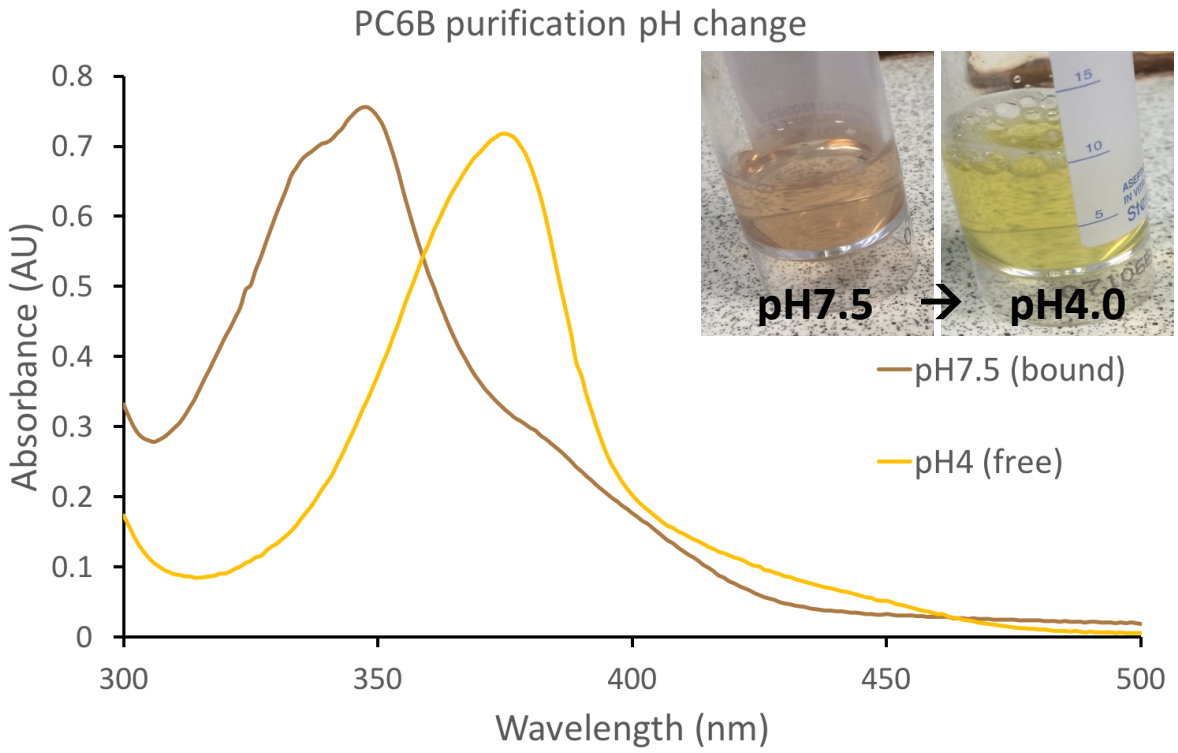


Figure 46: UV spectroscopy of PC6B purification and the effect when the pH is lowered to remove CobK protein.

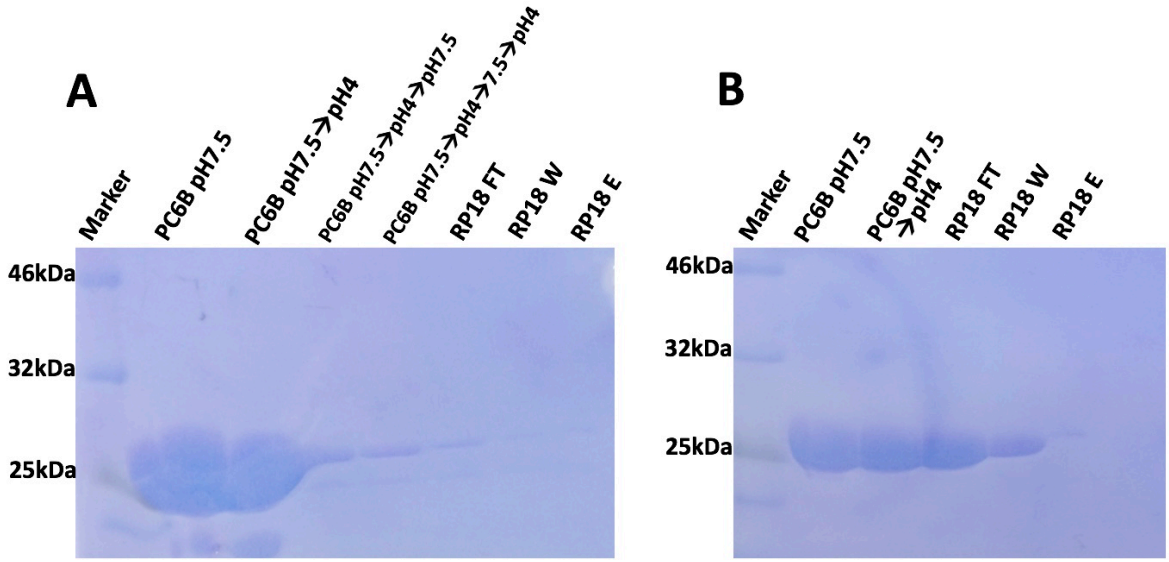


Figure 47: SDS PAGE gels of new (A) and old (B) methods to purify PC6B. A) Shows that an additional step changing the pH results in the removal of more CobK (~26kDa) earlier in the purification. B) Shows the old PC6B purification method where it can be seen a small amount of CobK remains in the final elution step from the RP18 column.

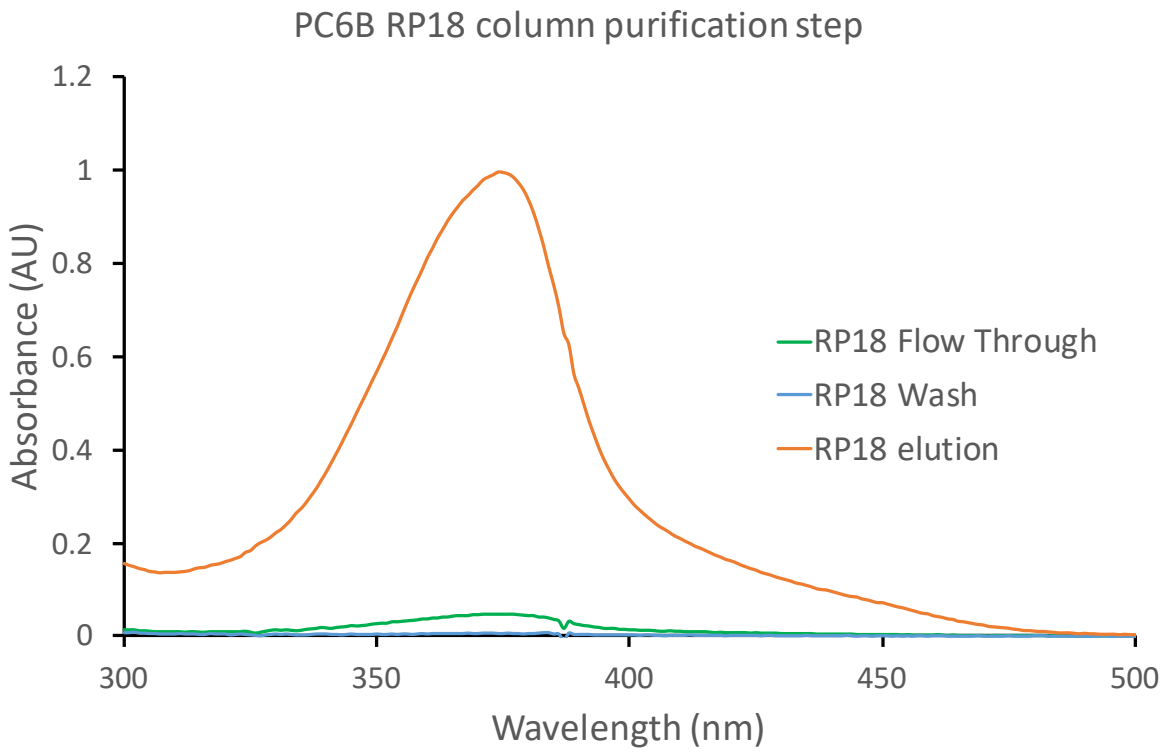


Figure 48: UV spectroscopy of PC6B RP18 purification step, showing free PC6B present in the elution step.

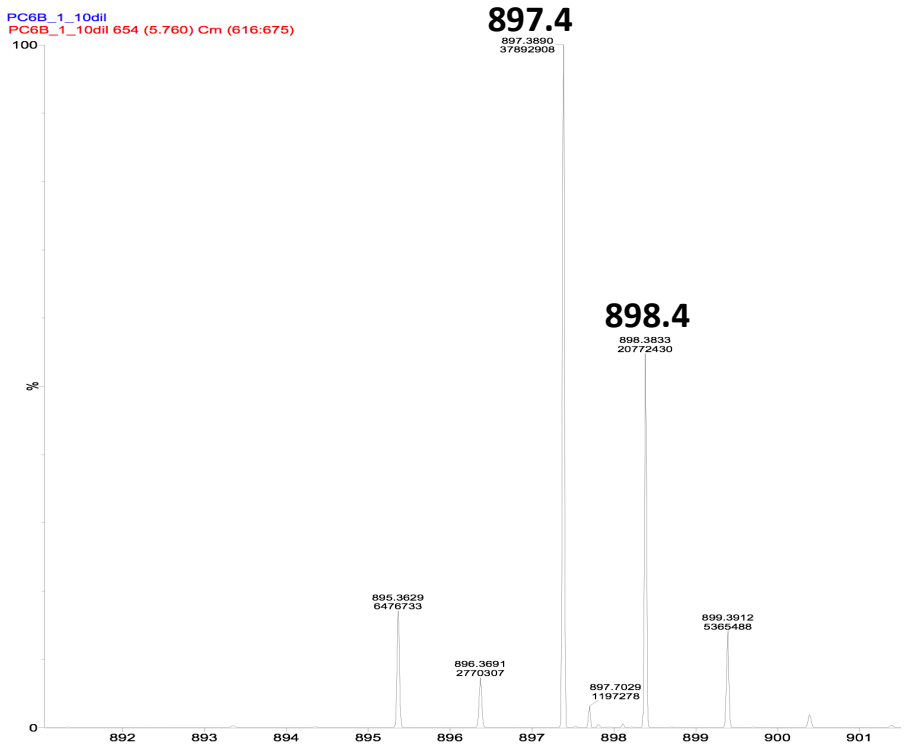


Figure 49: Mass spectroscopy of PC6B purification, showing the correct mass observed at 897.4.

4.2.2 PC6B wavelength scans under different buffer conditions

Initially CobK and CobL were purified in their respective buffers. However, studies involving both proteins CobK and CobL sometimes led to confusing wavelength scans relating to bound and free PC6B as recorded in section 4.2.1. As CobK was a more stable protein than CobL and therefore easier to purify, CobK could also be purified in CobL buffer during a protein purification, when both proteins were required for an experiment. Due to the misleading wavelength scans obtained the spectroscopy profile of PC6B was investigated by exploring the effects of pH and salt concentrations on purified PC6B, based on differences in the protein purification buffers used for CobK and CobL.

Initially the salt concentration was investigated as CobK is typically purified in 100mM NaCl pH7.5, while CobL is purified in 500mM NaCl pH8.0. A PC6B concentration of 30µM was diluted in varying concentrations of sodium chloride and a wavelength scan carried out to observe if varying the salt concentration altered the spectroscopy of PC6B. The results are shown in Figure 50, where it can generally be seen that increasing the salt concentration does not have an effect on the PC6B peak position (378nm), despite a general increase in the baseline of each absorbance reading.

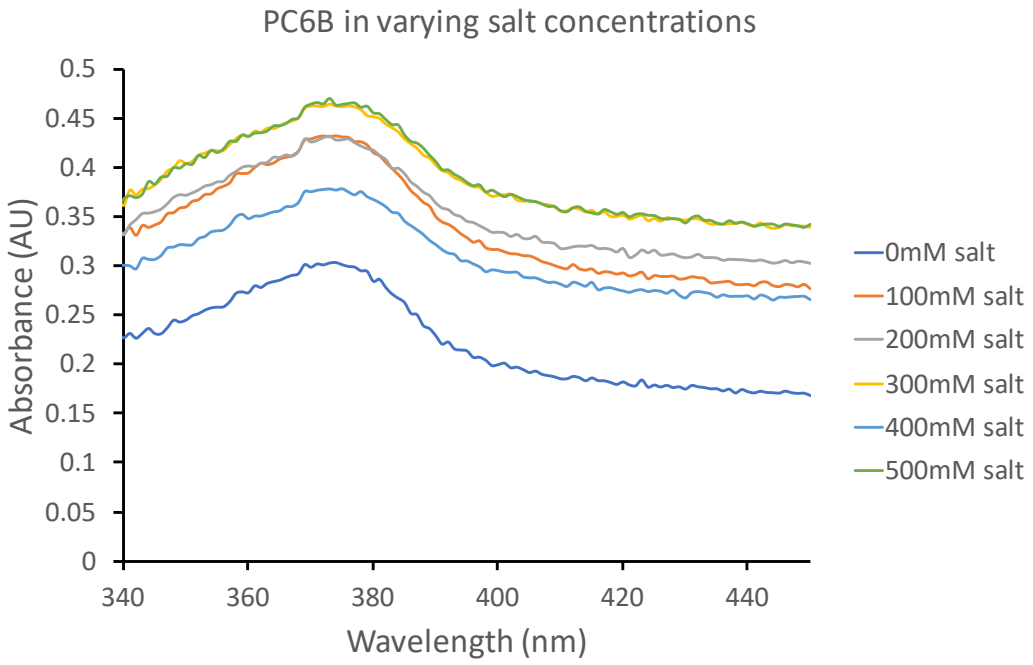


Figure 50: UV spectroscopy of PC6B in increasing sodium chloride concentrations in line with the protein purification buffers used in the laboratory.

However, the pH influenced the absorbance spectrum of PC6B as shown in Figure 51. There is a peak shift as pH is changed, with the maximum varying between 348nm and 378nm or a broader double peak between the two suggesting two to three different states of PC6B may exist. The yellow colour was visibly seen to decrease as the pH was increased. This change is most likely to be due to protonation of the nitrogen atoms inside the pyrrole groups, as this would affect the conjugation system of the tetrapyrrole. It is unlikely to be due to the carboxyl groups outside the tetrapyrrole as these are not directly conjugated to the ring structure. Finally, the brown colour observed in section 4.2.1 is not observed suggesting this colour is only present when bound to CobK, as observed after the nickel elution step during a purification, as shown in Figure 46.

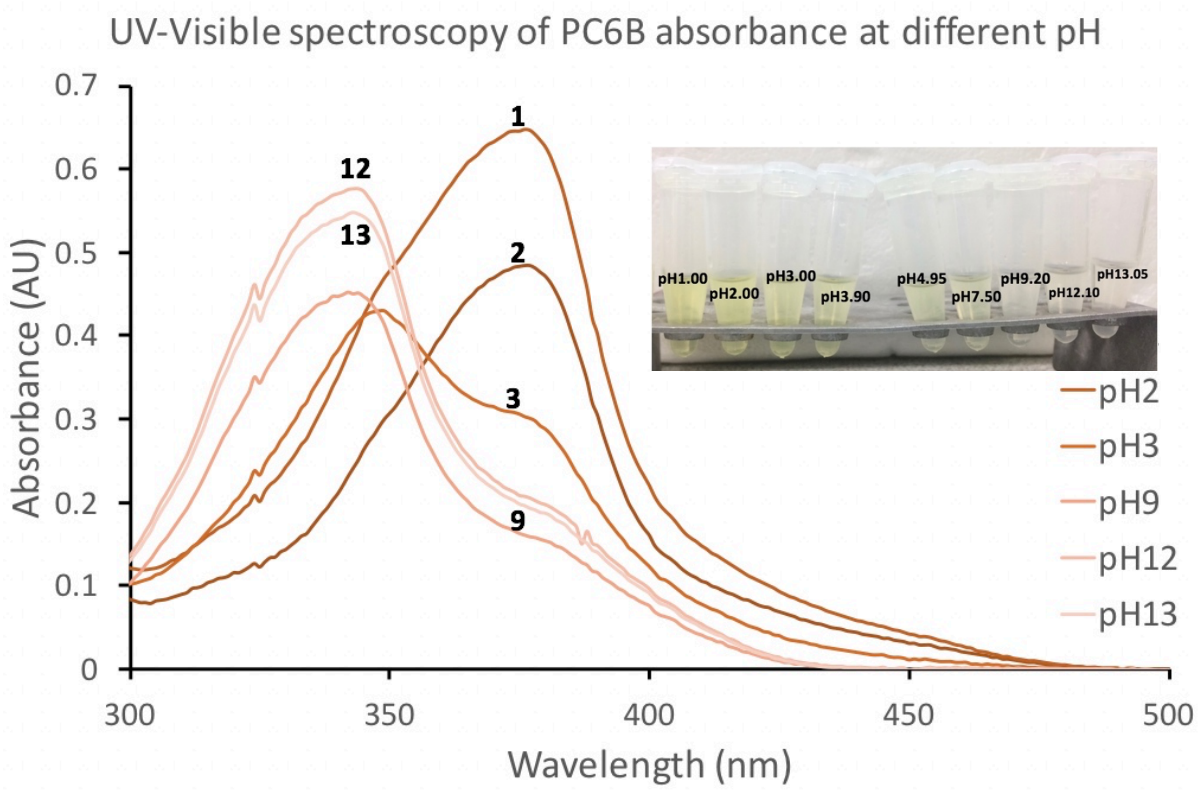


Figure 51: UV spectroscopy of purified free PC6B at different pH. It can also be visibly seen that the yellow colour decreases as the pH increases.

These results clearly show that pH has a profound influence on the absorbance spectrum of PC6B. This meant that all proteins needed to be purified in the same buffer to ensure pH changes did not influence the spectrum obtained. To avoid any further irregular PC6B wavelength scans,

concentrated stock solutions were subsequently prepared by weighing out a known mass of PC6B powder and dissolving this in distilled water to make stock solutions which were stored at -80^{0}C.

4.2.3 PC6B binding titration

The purification of PC6B via a CobK nickel column purification shows product binding activity is occurring. To further show that CobK is able to bind its product PC6B, a binding titration was carried out and measured using UV spectroscopy. A stock solution of 30µM PC6B was prepared in CobK buffer (pH7.5) and an initial wavelength scan carried out to show PC6B is in the free state which absorbs at 378nm. CobK both with and without NADPH present was titrated into this PC6B solution and a wavelength scan (280-450nm) taken after each addition. The results are shown in Figure 52 and show a clear peak shift from 378nm to 348nm over the course of the titration, showing PC6B absorbance changes from a free state to a CobK-bound state.

When the titration was carried out in the presence of CobK and NADPH there was a more distinct peak shift than when the titration was carried out only in the presence of CobK. An isosbestic point was observed around 360nm suggesting that the mixture contains two distinct states which are free PC6B and protein bound PC6B. This shows that NADPH assists in substrate (and product) binding and is most likely due to CobK loop interactions with both NADPH and PC6 allowing the loop to fully close over the active site.

NADPH does absorb with a peak maximum of 345nm, so a control titration was carried out where only NADPH was titrated into PC6B, as shown in Figure 52C. A slight peak shoulder is observed as the NADPH is titrated in, however the free PC6B peak maximum at 378nm is still clearly the distinct peak, showing that addition of NADPH at these low concentrations is clearly not enough to produce a false positive meaning the PC6B peak shift from 378nm to 348nm is due to CobK binding PC6B, with PC6B going from a free state to a bound state.

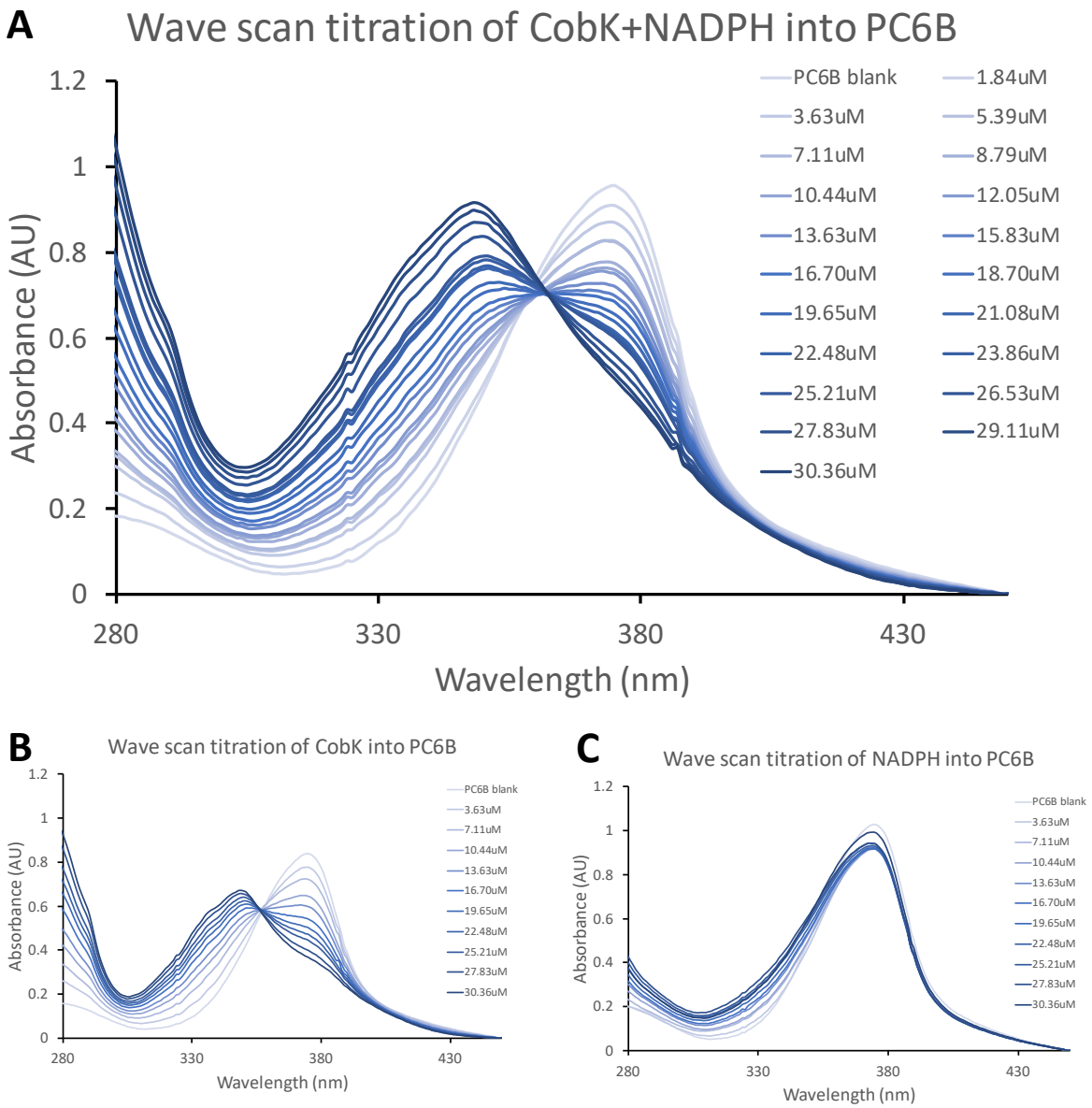


Figure 52: UV spectroscopy titrations of CobK and/or NADPH into free PC6B. A) Increasing additions of CobK+NADPH titrated into 30μM PC6B, where the free PC6B peak at 378nm shifts to 348nm. B) Increasing additions of CobK titrated into 30μM PC6B, where the free PC6B peak at 378nm shows a reduced absorbance maximum when the absorbance shifts to 348nm. C) Control titration where increasing NADPH was titrated into 30μM PC6B.

4.2.4 CobK thermal melt

To probe the product binding activity of CobK further, a thermal melt study was carried out in the presence and absence of NADPH and PC6B. Pure CobK protein was purified and each condition was repeated three times so an average could be taken. Each well contained 65µM CobK protein with a 5x excess ligand concentration used. A real time PCR thermal melt was carried out across a temperature range of 25-95^{0}C where the temperature was increased by 1^{0}C every minute between scans. The results are shown in Figure 53.

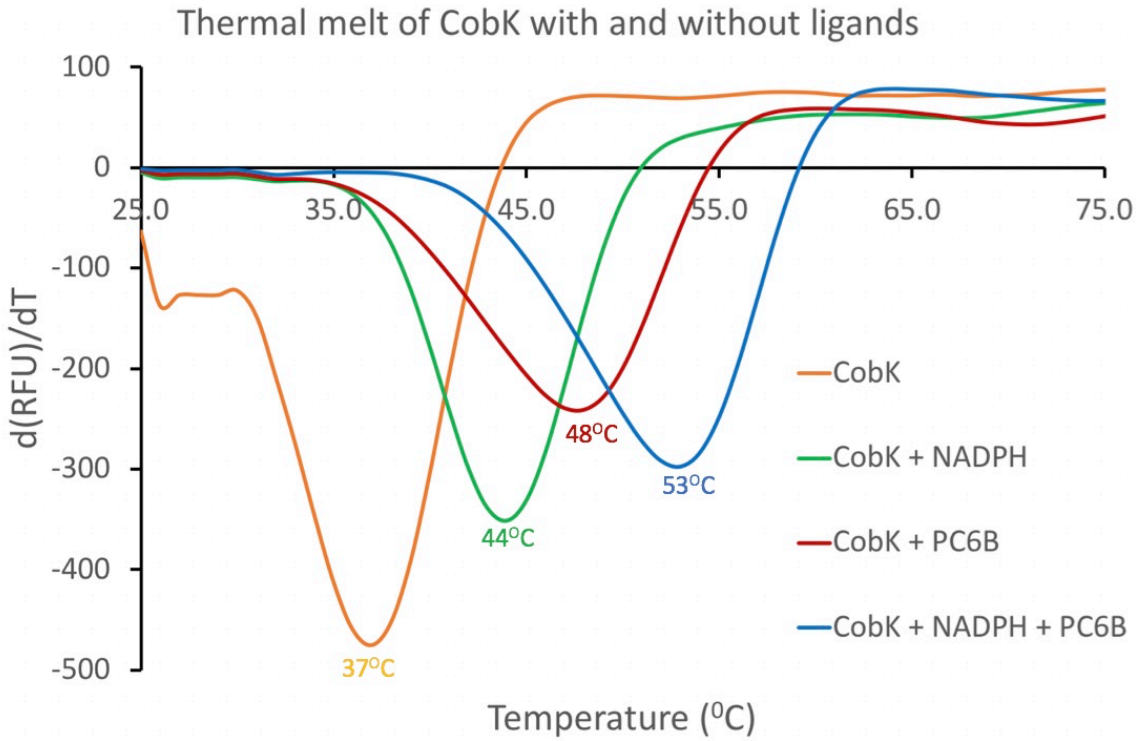


Figure 53: A real time PCR thermal melt experiment of CobK in the presence and absence of the ligands NADPH and PC6B. Upon the addition of these ligands an increase in T_m can be observed inferring an increase in protein stability. The melting temperature (T_m) can be calculated by recording the temperature at which the negative first derivative occurs and has been labelled on each curve.

CobK had a melting temperature (T_m) of 37^{0}C and the presence of NADPH and/or PC6B increased the T_m indicating increased protein stability when these ligands are bound. The addition of NADPH increased the T_m to 44^{0}C while the addition of PC6B increased the T_m to 48^{0}C. However, the biggest increase in T_m came when both NADPH and PC6B were added together to CobK increasing the T_m to 53^{0}C. This shows that the addition of a ligand to CobK increases the protein's stability. The increase

in T_m when both NADPH and PC6B are present further supports tight product binding activity of CobK and the addition of two ligands, NADPH and PC6B, together causes a large increase in stability which is most likely due to the loop closure around the active site making the protein more rigid.

4.2.5 NMR of CobK product binding activity

A 2D NMR HSQC experiment was carried out to show CobK binding NADPH and PC6B. The protein was produced by growing *E.coli* in minimal media supplemented with N^{15} labelled ammonium chloride to produce labelled CobK protein which was purified following the standard protein purification described in section 2.7.2.1 down a nickel column followed by a size exclusion column. An SDS-PAGE of the successful protein purification is shown in Figure 54. As expected, the yields were less than those normally observed, as in section 3.2.4, due to the protein being produced in minimal media. There was also a small amount of unlabelled CobK present which was likely to be due to the N^{15} labelling protocol followed, in which initially the cells were grown in LB media to produce all the cell machinery required for protein synthesis before being substituted with minimal media containing N^{15}. Additional proteins can be seen in the SDS gel, possibly due to stress proteins produced by the *E. coli* resulted in a very broad UV absorbance during size exclusion chromatography.

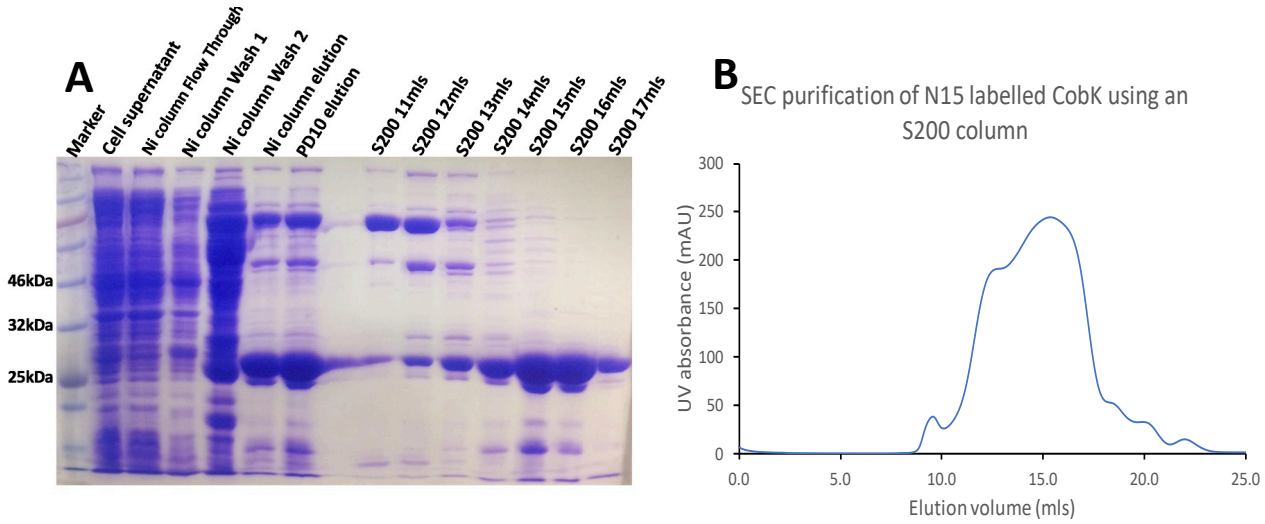


Figure 54: Purification of ^{15}N labelled CobK (~26kDa). A) SDS-PAGE analysis of the protein purification process. B) UV profile of the S200 size exclusion column elution where the peak is from the fractions 11 to 17mls.

Once purified the N^{15} labelled CobK was concentrated to a concentration greater than 0.2mM for the NMR experiment. The NMR spectra are shown in Figure 55and show native CobK, CobK + NADPH, CobK + NADPH and PC6B. All three spectra are different highlighting that CobK is interacting with the ligands. Despite the amino acids of the protein not being assigned it does suggest the ligands are interacting with CobK. The changes observed in the spectra are most likely due to two halves of CobK being flexible around the central active site, which acts similar to a hinge when the loop region closes over it. Upon the addition of the ligands NADPH and PC6B peaks shifts are observed showing the protein is undergoing a conformation change upon binding.

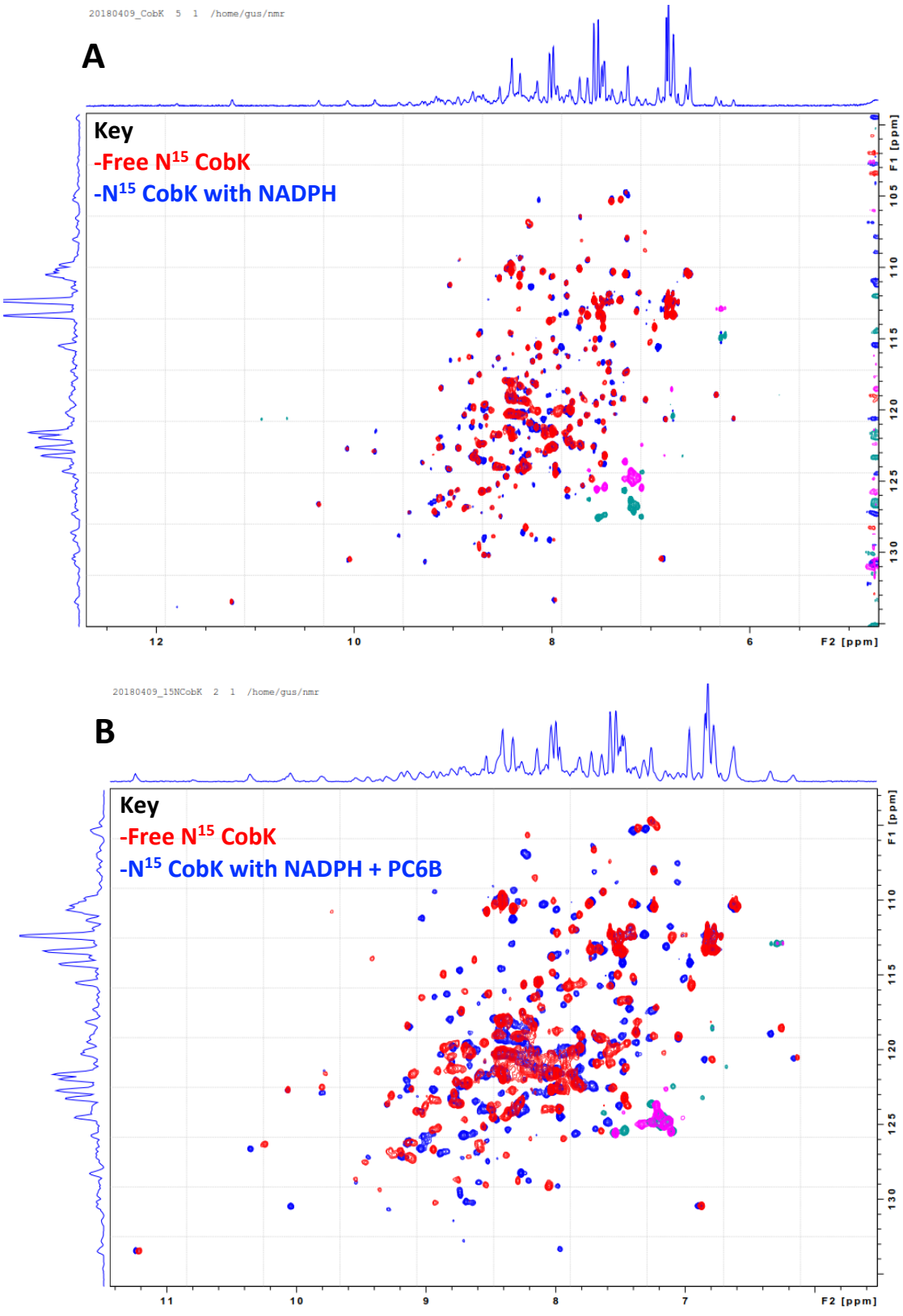


Figure 55: HSQC NMR of N^{15} labelled CobK with the ligands NADPH and PC6B. A change in the spectra can be seen upon addition of ligands to the protein. A) A comparison of free CobK in red and CobK with NADPH present in blue. B) A comparison of free CobK in red and CobK with NADPH and PC6B present in blue.

4.2.6 Isothermal titration calorimetry (ITC)

A final experiment to probe the product binding activity of PC6B to CobK was ITC. Initial trial runs (data not shown) were undertaken at 20^0C with 13 injections of 3µl and included titrating; PC6B into CobK buffer, PC6B into CobK and PC6B into CobK with NADPH present (also at 30µM). The first titration of PC6B into buffer did not yield any change and confirmed no buffer mismatch. The second titration trial of PC6B into CobK did result in an endothermic signal but the change was only small. The final titration of PC6B into CobK and NADPH again resulted in an endothermic change with a larger signal, but saturation occurred too quickly. For future titrations it was concluded that PC6B into CobK + NADPH was the experiment to be pursued. As ligand saturation occurred too quickly the number of injections was increased to 19 with a volume of 2µl, and the concentration of protein was increased to 45µM to increase the signal and the temperature lowered to 15^0C given the reaction observed was endothermic.

The final titration was repeated three times and all run at 15^0C which are shown in Figure 56. The CobK concentration used in the cell was 45µM and PC6B concentration in the syringe was kept at 300µM. The dissociation constant (K_d) measured was 2.97µM with an error of $7.61x10^{-7}$ and a stoichiometry around 0.85. The binding of PC6B was endothermic giving an enthalpy change of 2.7 kcal/mol. This K_d value is at the higher limit of moderate binding affinity typically quoted between millimolar to micromolar ranges[112]. Stoichiometry of around 0.85 suggests the PC6B concentration is underestimated by 15% or concentration errors in ligand and enzyme combined to produce a 15% error. 3µM is significant binding, but not as tight as drug-receptor interactions which may be 1000 times tighter in the nanomolar ranges[112].

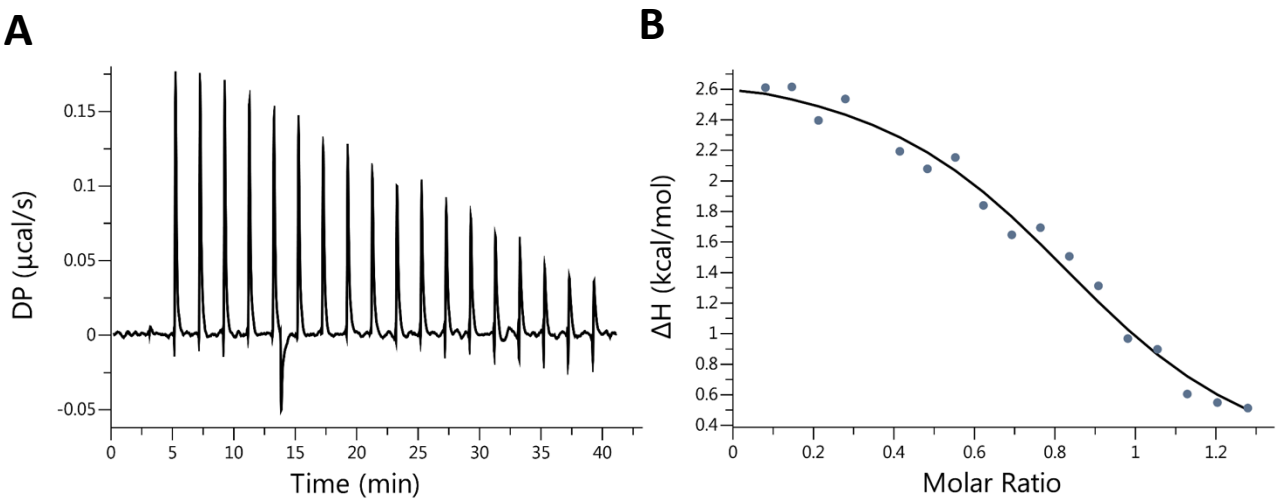


Figure 56: ITC titration of PC6B titrated into CobK. A) Shows the thermal heat signature observed during each injection. B) Shows the change in enthalpy measured throughout the experiment.

4.3 Probing protein-protein interactions between CobK and CobL

4.3.1 Native gel studies

Some experiments were initially carried out to probe whether CobK and CobL interact. The first was a native PAGE experiment where CobK and CobL were mixed with ligands to see if an interaction could be observed. Figure 57 shows the native page gel of CobK mixed with CobL with the ligands NADPH, PC6B and SAH present. As CobL forms a tetramer with four active sites, it is possible four CobK proteins may be able to bind to the C-terminal domains. As the binding ratio was not known an increase in the ratio of CobK:CobL was made across the gel exploring up to 4x CobK: 1x CobL.

In the native PAGE gel in Figure 57 SAH was used in replacement of SAM to prevent any PC6B turnover and hopefully encourage or prolong a protein interaction. As the concentration of CobK increases across the gel its band becomes more apparent when mixed with CobL. The emergence of a new band or disappearance of existent CobK and CobL protein bands does not occur, when compared to the control lanes suggesting there is no CobK-CobL interaction.

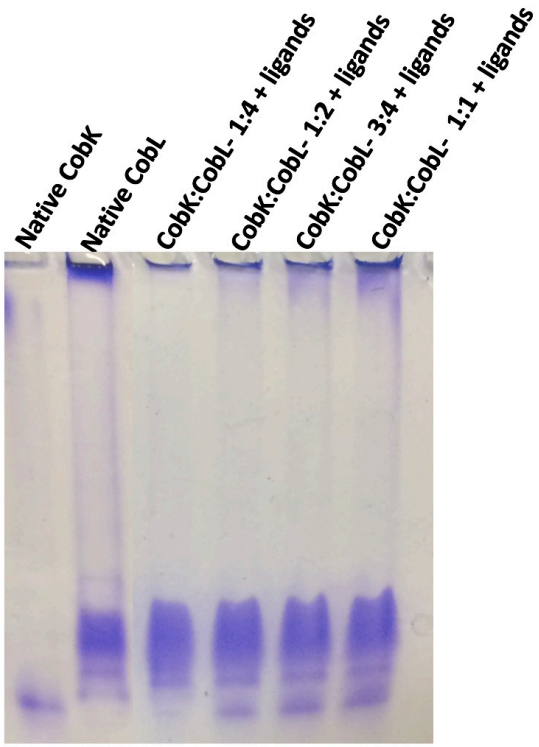


Figure 57: Native PAGE gel of CobK mixed at increasing ratios to CobL with the ligands NADPH, PC6B and SAH present.

4.3.2 Nickel column pull down

A nickel column pull down experiment was attempted to see if CobK and CobL interacted. This involved removing the Histidine tag from one of the proteins via the TEV cleavage site using thrombin. The thrombin cleavage was trialled on both proteins at different temperatures shown in Figure 58. The trial was successful with both proteins where successful cleavage was shown by a smaller molecular weight protein band on an SDS-PAGE gel. Both protein gels show that protein treatment with thrombin at 4^0C overnight was not enough for complete cleavage and 37^0C overnight was too high and caused a high activity of unspecific protein cleavage, destroying the proteins in the process. The best CobK thrombin cleavage occurred at 16^0C and 30^0C left overnight, while the best CobL thrombin cleavage occurred at 30^0C overnight.

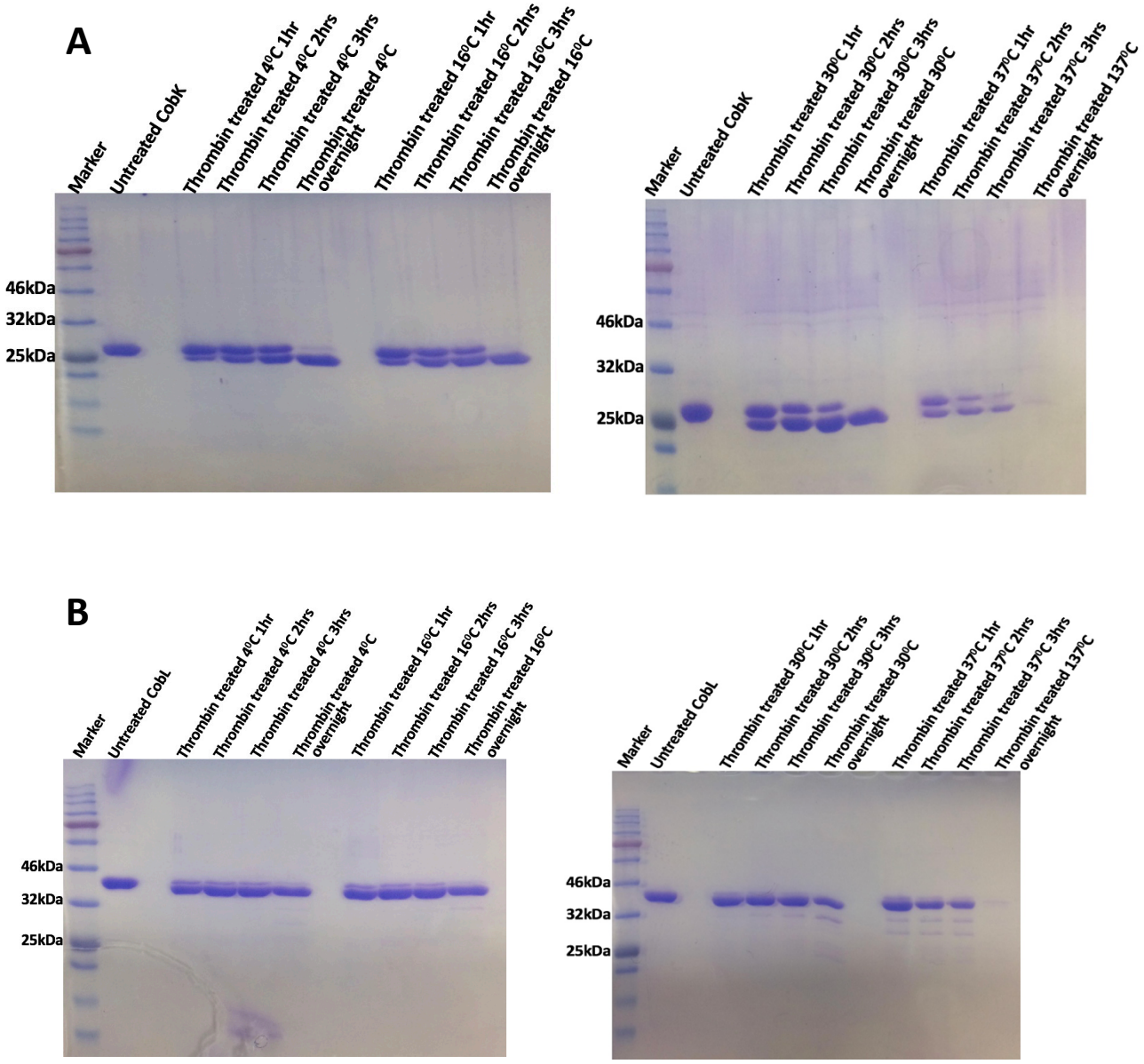


Figure 58: Thrombin TEV cleavage temperature trials on the proteins CobK ~26kDa and CobL ~42kDa. A) Shows the temperature trials of thrombin activity on CobK at 4^0C, 16^0C, 30^0C and 37^0C after one hour, two hours, three hours and left overnight. B) Shows the temperature trial conditions as in (A) but on CobL protein. As the temperature and time increases the improvement of thrombin activity in removing the Histidine tag can be observed by the reduction in protein mass. In both trials 16^0C overnight is the most successful, while 37^0C is too active and when left for too long destroys the protein via unspecific activity.

For the nickel column pull down experiment only one protein needed to have its histidine tag removed. CobK was selected and thrombin treated overnight. The sample was then passed through a nickel column, where the cleaved protein was collected from the nickel column wash step. 100μM Cleaved CobK was then mixed with 100μM CobL and ligands SAH and PC6B in a 5x excess. The mixture was then passed through a nickel column where the flow through, wash and elution steps were collected. The samples were run on an SDS-PAGE gel and the results are shown in Figure 59. The results show that the thrombin cleavage was successful, and protein isolated in the wash step. When cleaved CobK was mixed with CobL and ligands, if there was an interaction both would be present in the final elution step. However cleaved CobK is shown to be present in the wash step, with none present in the elution so there is no evidence from the pull down to suggest a CobK-CobL interaction.

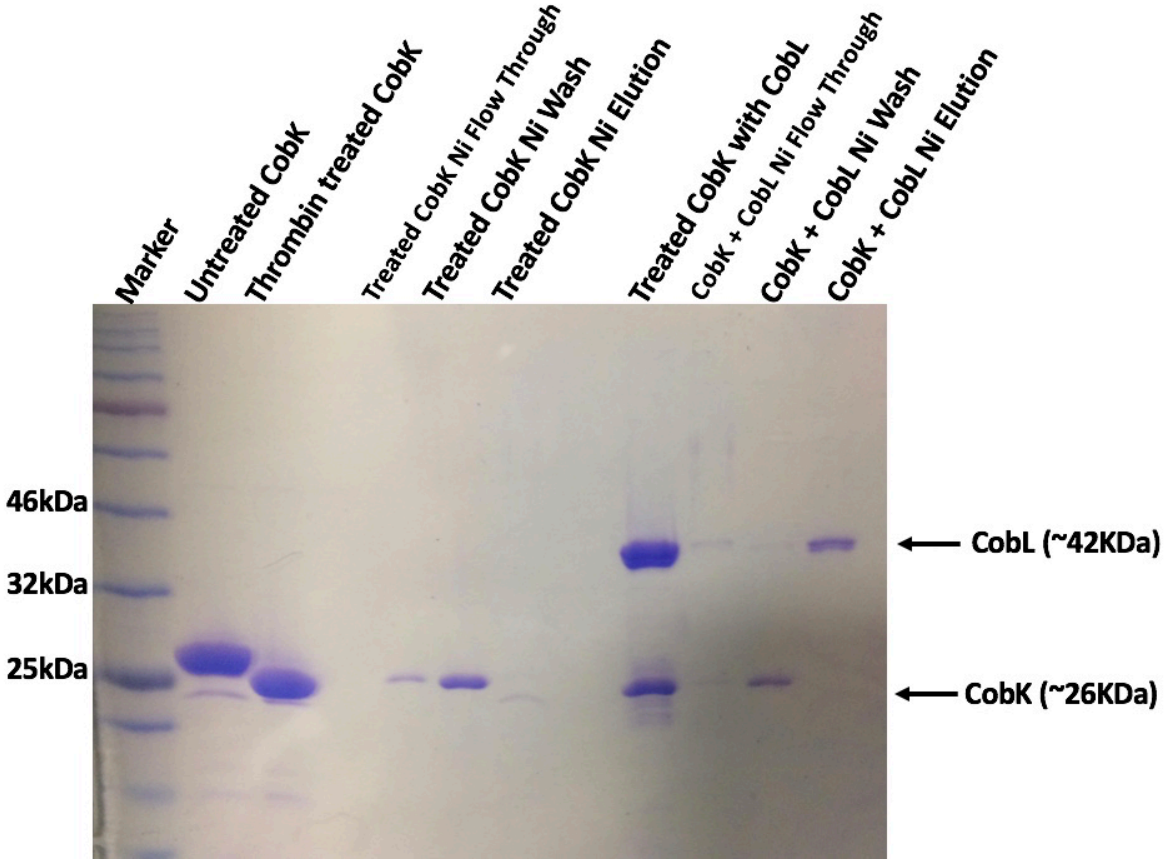


Figure 59: Nickel column pull down experiment between thrombin treated CobK ~26kDa and Histidine tagged CobL ~42kDa.

4.3.3 Size exclusion chromatography

A separate pull-down experiment was attempted using a size exclusion column to observe a potential interaction between CobK and CobL by way of monitoring if an earlier elution peak was due to the formation of a larger protein complex. The protein elution profiles are shown in Figure 60, where CobL is observed eluting the S200 column after 14mls, while CobK is observed eluting after 17mls. CobL comes off the column earlier as it is a larger protein with a molecular mass of ~42kDa (tetramer ~168kDa) so it is not retained for as long as CobK which has a molecular mass of ~26kDa.

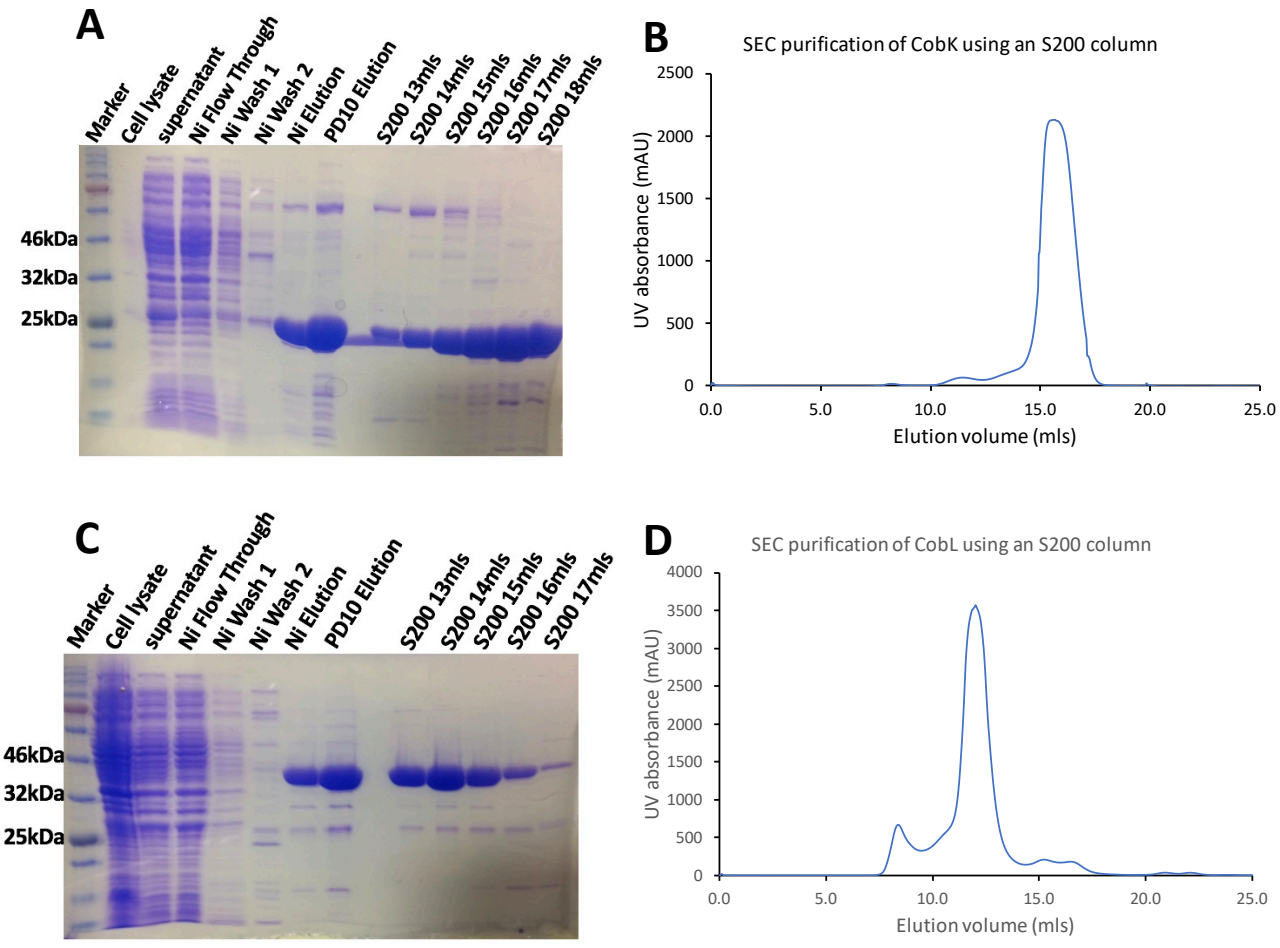


Figure 60: Purification of CobK (~26kDa) and CobL (~42kDa). A) SDS-PAGE analysis of CobK protein purification process. B) UV profile of the S200 size exclusion column elution of CobK where the peak is from the fractions 13 to 18mls. C) SDS-PAGE analysis of CobL protein purification process. B) UV profile of the S200 size exclusion column elution of CobL where the peak is from the fractions 13 to 17mls

When the two proteins CobK and CobL were mixed in equal concentrations and injected on to an S200 size exclusion column, two elution peaks were observed at the volumes CobL and CobK normally elute and this was confirmed using an SDS-PAGE gel shown in Figure 61A. The gel did not indicate any interaction as each elution aliquot only contained one protein. When the experiment was repeated in the presence of the ligands NADPH, SAH and PC6B in a 5x excess an additional elution peak was observed at 20ml, Figure 61B, but an SDS-PAGE confirmed this was CobK only. This was thought to occur due to the presence of the ligands NADPH and PC6B binding to CobK, making the protein more rigid upon product binding resulting in a smaller surface area and longer elution time. This was confirmed by injecting a separate sample of CobK with ligands back on to the size exclusion column and elution volumes compared in Figure 61C, which again confirms this additional peak is CobK.

The crystal structures of native and ligand bound CobK were overlaid to show that there is a change in protein conformation. Native CobK is more flexible around its central active site, where the two halves have more freedom to act independently resulting in a faster elution time due to the protein flexibility occupying a larger volume on the column. This conclusion is also in agreement with the thermal melt study in section 4.3.4, where ligand binding increases CobK stability and a conformation change observed by overlaying the native and ligand bound crystal structures of CobK highlighted in Figure 61D.

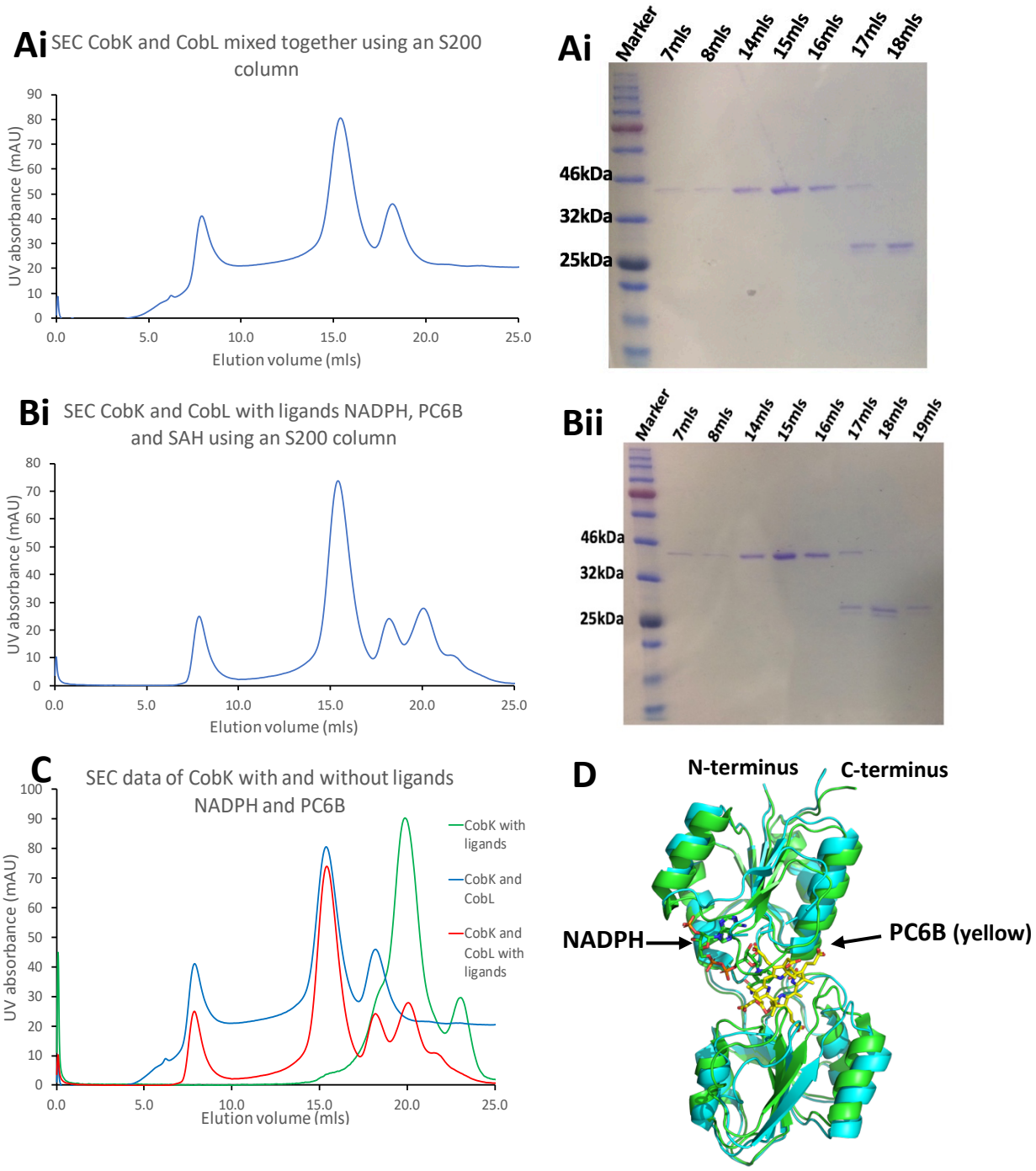


Figure 61: SEC pull data of CobK ~26kDa and CobL ~42kDa to see if they interact accompanied with SDS PAGE gels of the elution volume aliquots. A) Shows CobK and CobL injected together at equal concentrations. B) Shows CobK and CobL injected together with ligands NADPH, PC6B and SAH. C) Shows the different elution volumes of CobK in the presence and absence of the ligands NADPH and PC6B. D) The crystal structures of CobK (blue) alignment with CobK + NADPH & PC6B (green) to show an overall change in protein conformation. Figure D made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

4.3.4 Thermal melt studies of CobK and CobL

A thermal melt of 200µM CobK and CobL was undertaken in the presence of 20x ligands NADPH, PC6B and SAH or SAM, Figure 62. The melting temperature (T_m) measured with just CobK and CobL was 44°C, which was an increase on the previously reported T_m of CobK of 37°C and CobL of 38°C. In the presence of NADPH, PC6B and SAM the T_m increased to 48°C, whereas in the presence of NADPH, PC6B and SAH the T_m increased to 51°C. These T_m values do not quite align with previous thermal studies of CobK and CobL, possibly indicating a protein-protein complex is forming.

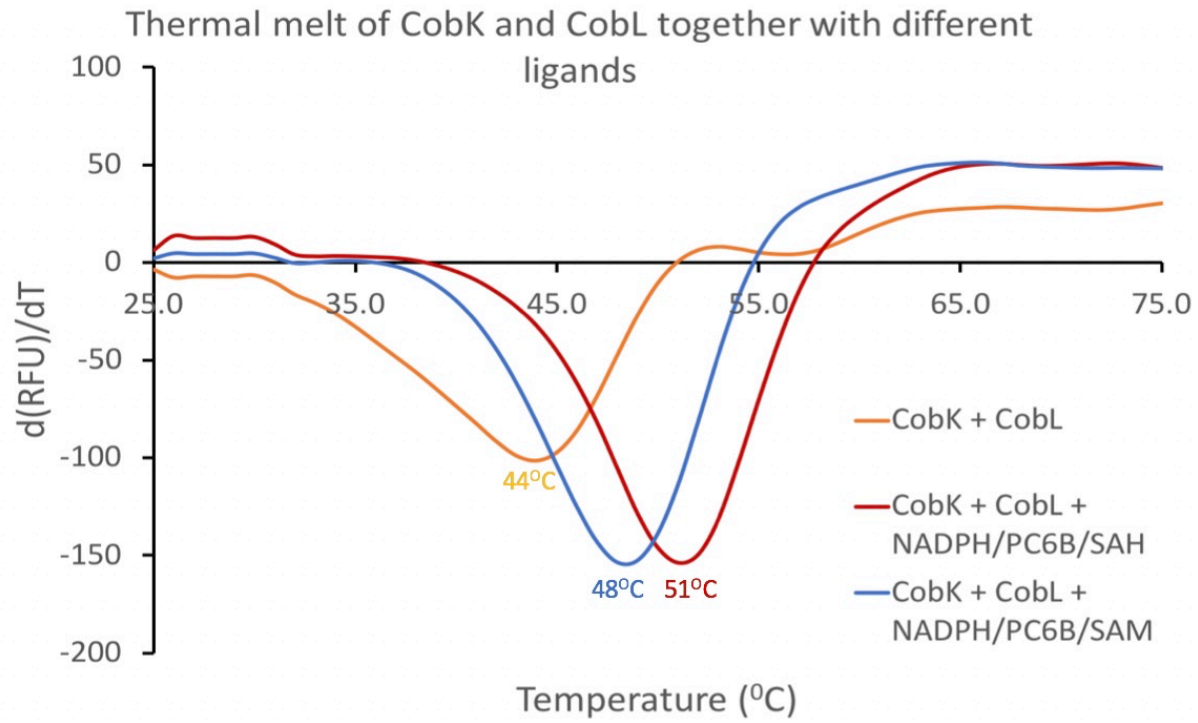


Figure 62: A real time PCR thermal melt shift of CobK and CobL together in the presence and absence of ligands which included combinations of NADPH, PC6B and SAH or SAM. The melting temperature (T_m) can be calculated by recording the temperature at which the negative first derivative occurs and has been labelled on each curve.

4.3.5 Does SAH cause PC6B release?

With CobK having been shown to bind its product PC6B, this suggested that it might be for protection of intermediates in the biosynthetic pathway. Class I amidotransferases, such as CobB, have been observed to only degrade glutamine when ATP and HBA are present to prevent the wasteful loss of generated ammonia. Perhaps CobK works in a similar fashion. It was noted that SAM/SAH both contain an adenine functional group linked to a ribose similar to NADPH. As SAM is the ligand required in the next biosynthetic step perhaps it works as a trigger molecule to open the loop to facilitate PC6B release. This was investigated using UV spectroscopy to establish if PC6B was released.

A PC6B purification was carried out up to the nickel column elution step to provide a sample of PC6B bound to CobK. The protein concentration of this sample was measured so that increasing concentrations of both SAM and SAH could be titrated in to CobK protein with PC6B bound. The ratio of SAM/SAH to CobK protein was increased from 1:1 to 50:1 where the CobK concentration used was 52µM. The results are shown in Figure 63and show that increasing the respective concentration of SAH/SAM with respect to CobK with PC6B bound does not stimulate the release of PC6B as no peak shift is observed, with only the PC6B bound peak present at 348nm.

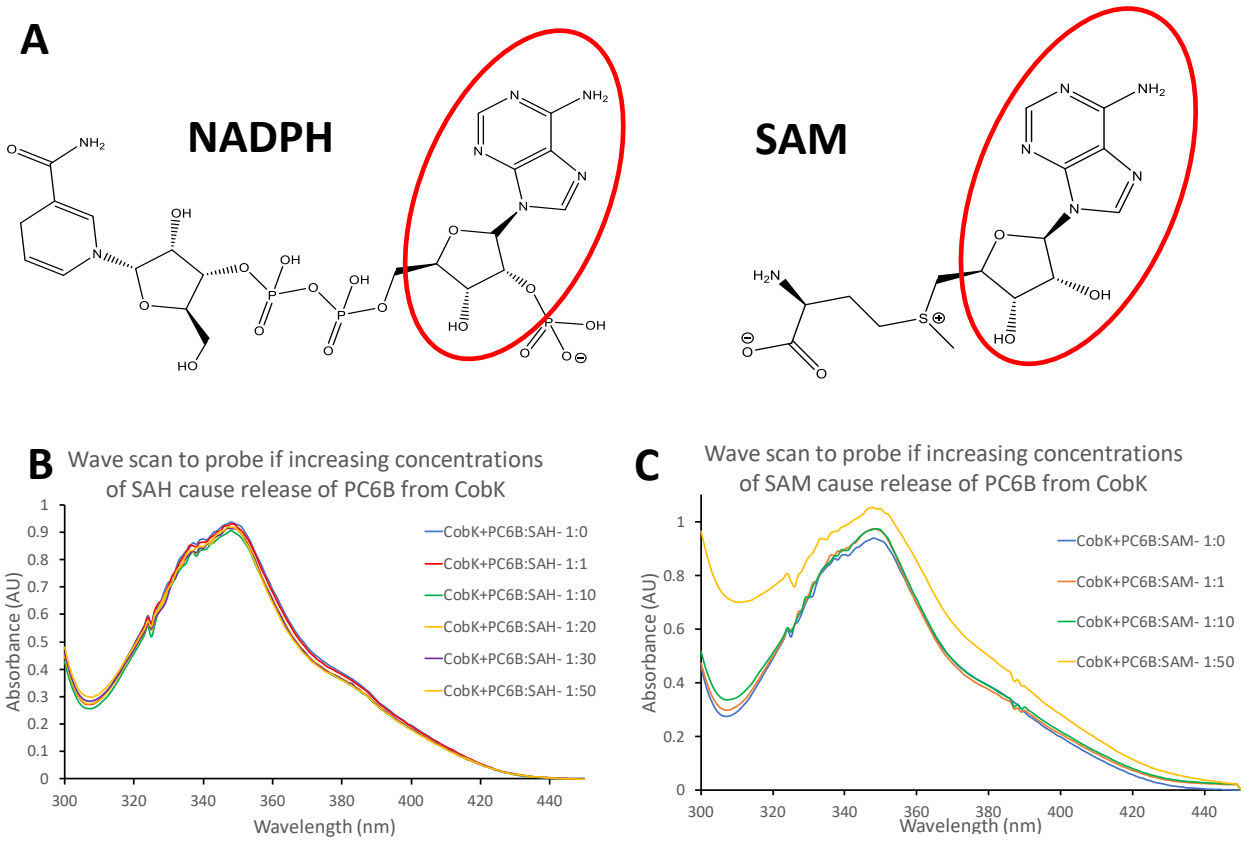


Figure 63: A) Shows a comparison of the chemical structures of NADPH and SAM. B) and C) Show a UV titration of the addition of SAH and SAM respectively, where both are titrated into CobK with PC6B bound to see if they trigger the release of PC6B which absorbs at 348nm when bound and 378nm when free in solution.

4.3.6 Enzyme kinetics

To explore if CobK and CobL interact the turnover of PC6B in the presence and absence of CobK was investigated to see if a change in the rate of reaction was observed. PC6B has an absorption maximum of 348nm when in the presence of CobK whereas PC8 shows a decrease and shift with an absorbance maximum of 390nm. Due to the stronger absorption signal at 348nm this was chosen as the wavelength to measure absorption, as shown in Figure 64.

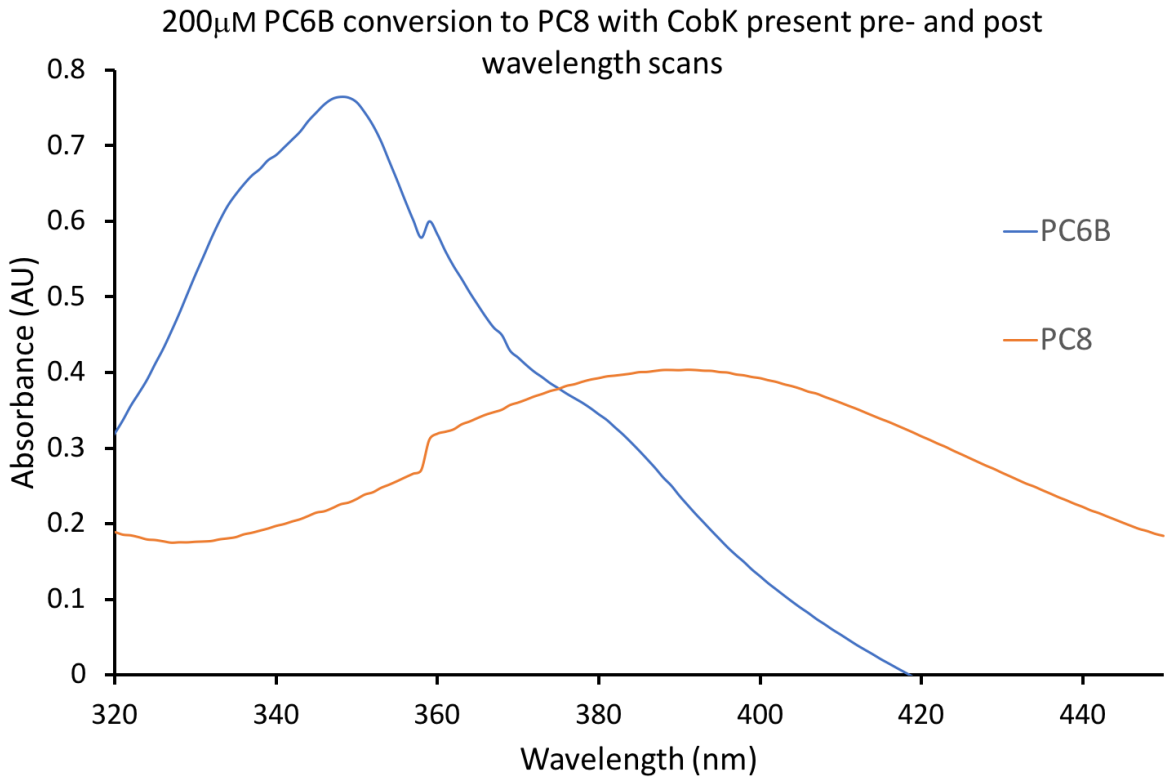
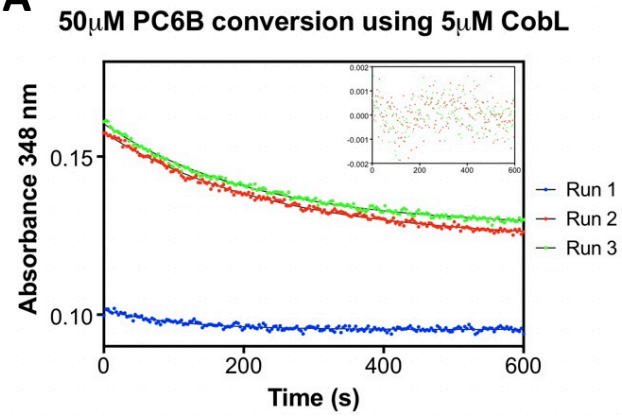
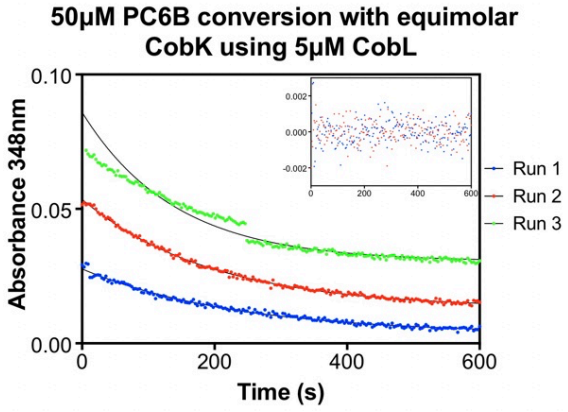


Figure 64: A PC6B and PC8 absorbance of kinetic studies to observe the turnover of PC6B by CobL in the presence of CobK. PC6B has an absorbance maximum of 438nm and PC8 has an absorbance maximum of 390nm.

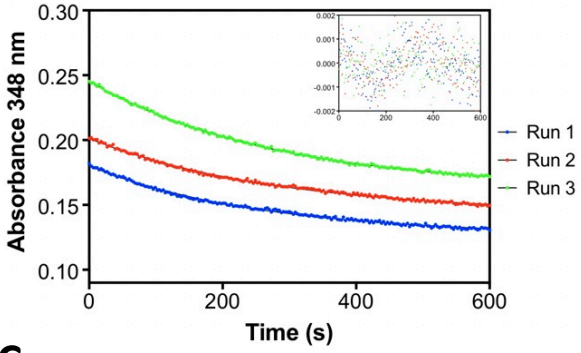
The turnover of PC6B was measured at 50μM intervals between 50μM-400μM in the presence of 5μM CobL, where the reaction was initiated by the addition of 1mM SAM and A_{348nm} was measured every three seconds over ten minutes. The second investigation of PC6B turnover was as above but with the addition of equimolar concentrations of CobK & NADPH to PC6B to ensure product binding. These PC6B conversions are shown in Figure 65where three runs were recorded for each concentration and the data was fitted to Equation 4.

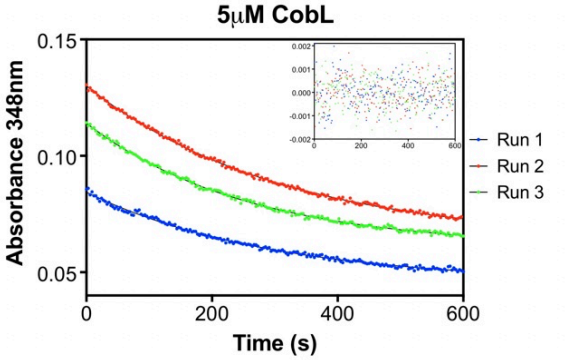
A


50μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

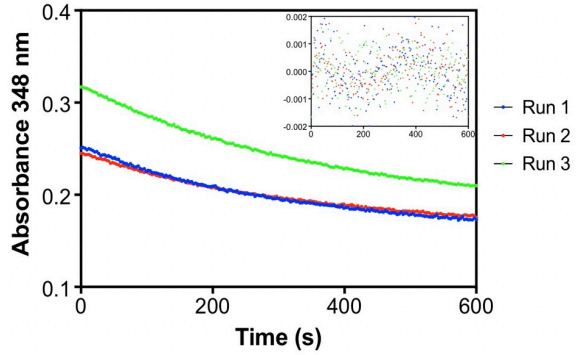

50μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

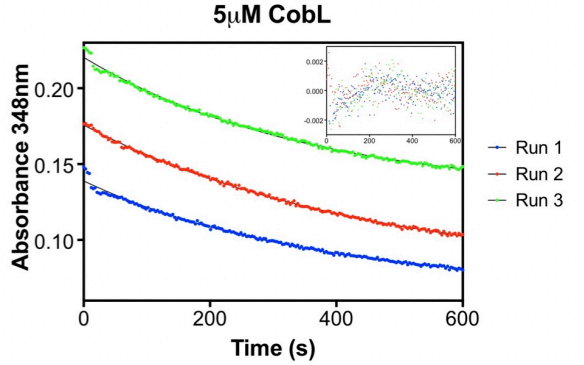
B


100μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

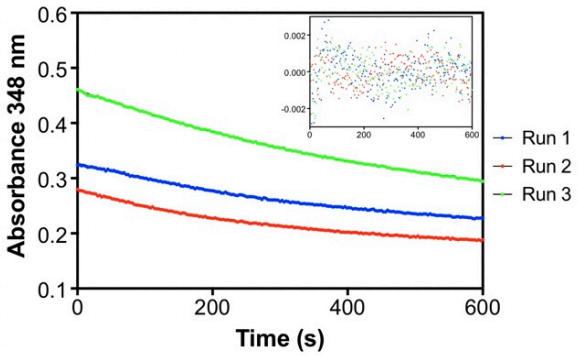

100μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

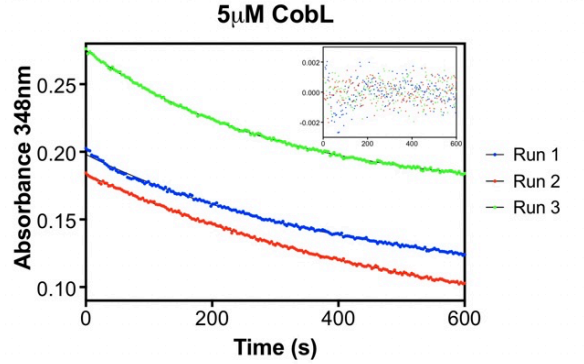
C


150μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

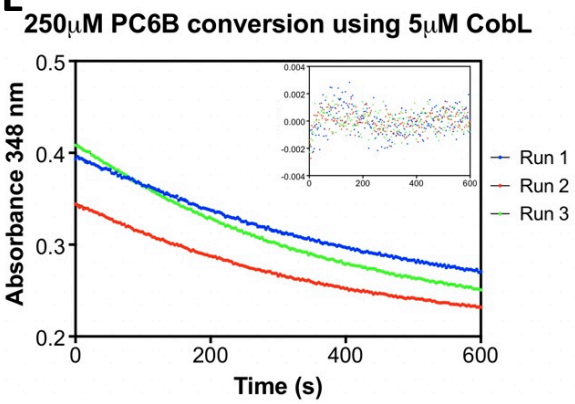

150μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

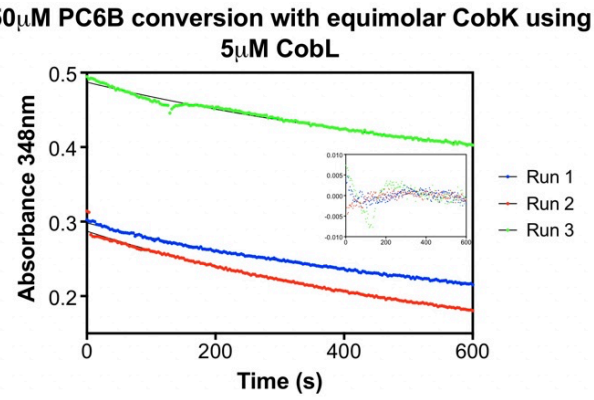
D


200μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

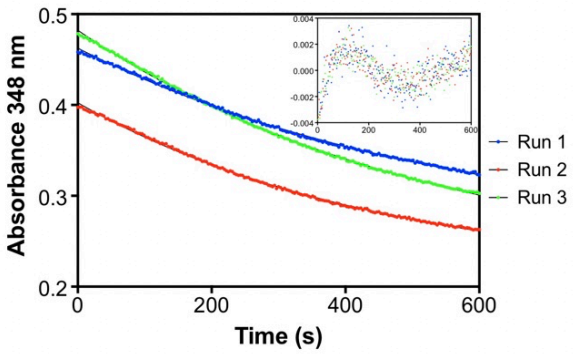

200μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

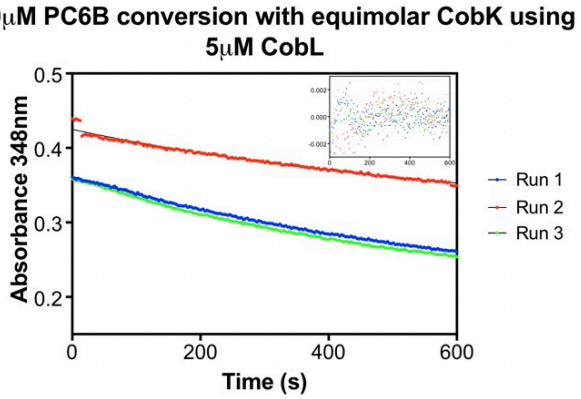
E


250μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

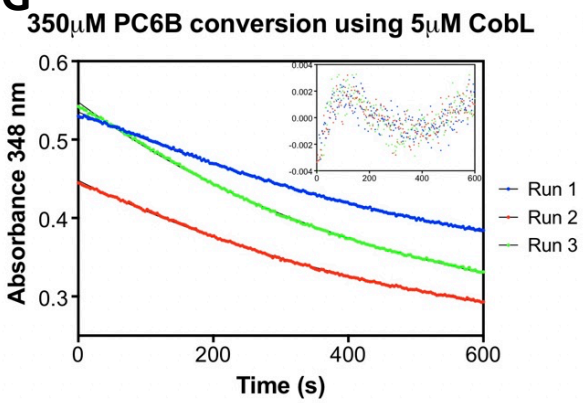

250μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

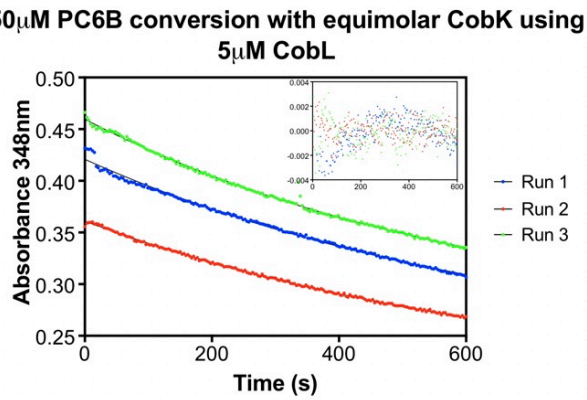
F


300μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

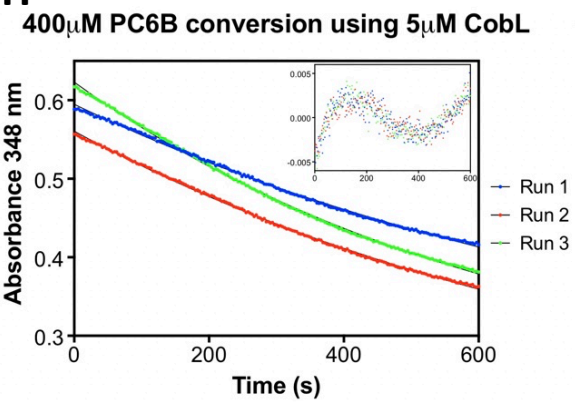

300μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

G


350μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3


350μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

H


400μM PC6B conversion using 5μM CobL
Absorbance 348 nm
Time (s)
Run 1
Run 2
Run 3

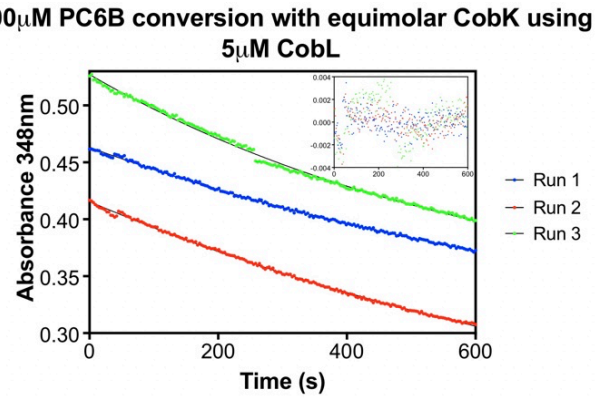

400μM PC6B conversion with equimolar CobK using 5μM CobL
Absorbance 348nm
Time (s)
Run 1
Run 2
Run 3

Figure 65: Kinetics data collected for varying concentrations of PC6B turnover in the presence and absence of equimolar CobK and NADPH using 5μM CobL and 1mM SAM to begin the reaction. (A) 50μM PC6B (B) 100μM PC6B (C) 150μM PC6B (D) 200μM PC6B (E) 250μM PC6B (F)300μM PC6B (G) 350μM PC6B (H) 400μM PC6B.

From Equation 4 a rate constant k was obtained, and this allowed an initial velocity to be calculated using Equation 5. Based on the residual plot scatter appropriate runs were selected so that an average velocity could be plotted against PC6B concentration in Figure 66. These data were then fitted using the Michaelis-Menten equation, Equation 6, where the error quoted was ± standard deviation between the initial rates measured for each run. This allowed a reciprocal Lineweaver-Burk plot to be plotted in Figure 66B allowing the V_{max}, K_M and k_{cat} to be calculated and are quoted in Table 10. Both the V_{max} and K_m were provided by Prism analysis of the Michaelis-Menten graph but were crossed checked via the Lineweaver-Burk plot where the y-axis intercept is $1/V_{max}$ and the x-axis intercept is $1/K_m$ in which they were almost identical.

Equation 4

$$y = Y_0\, exp^{-kt} + C$$

Where: y= absorbance (AU), Y_0= absorbance at 0s (AU), k= rate constant ($moldm^{-3}s^{-1}$), t=time (s), C= y-intercept

Equation 5

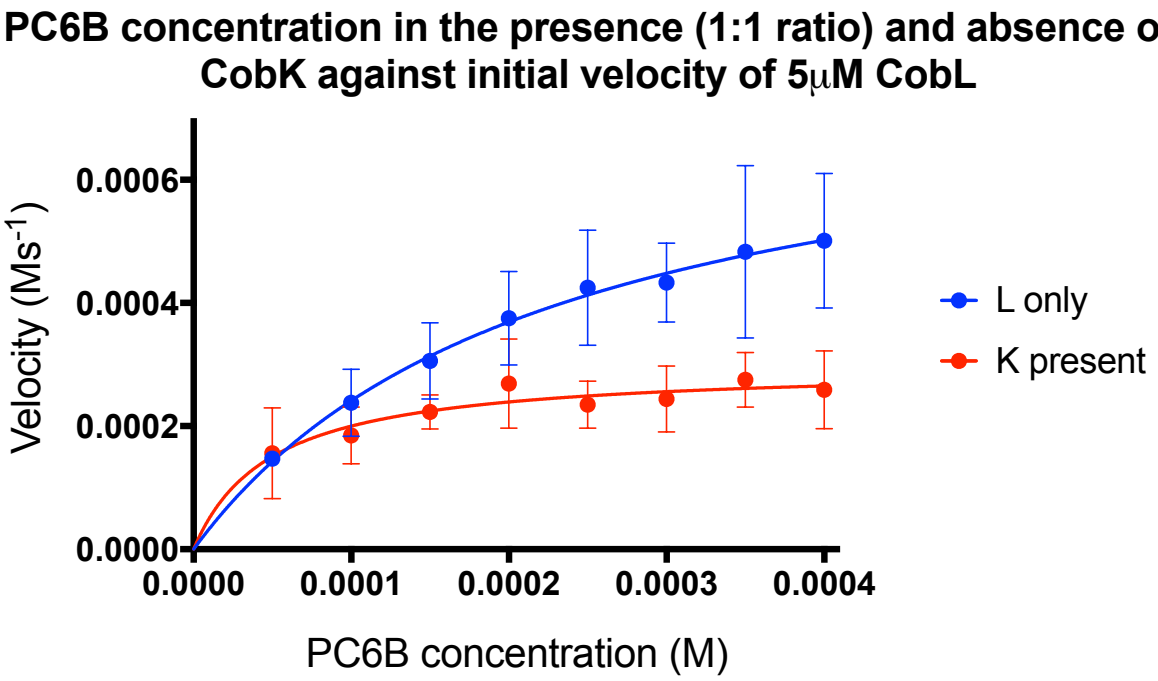
$$Initial\ rate = k\, Y_0$$

Where: k= rate constant ($moldm^{-3}s^{-1}$), Y_0= absorbance at 0s (AU) and units of Initial rate are $moldm^{-3}s^{-1}$.

Equation 6

$$V = \frac{k_{cat}[E][S]}{k_M\,[S]}$$

Where V= enzyme velocity ($moldm^{-3}s^{-1}$), k_{cat}= enzyme catalytic turnover (s^{-1}), [E]= enzyme concentration ($moldm^{-3}$), [S]=substrate concentration ($moldm^{-3}$), K_M= Michaelis constant ($mol^{-1}dm^3$).



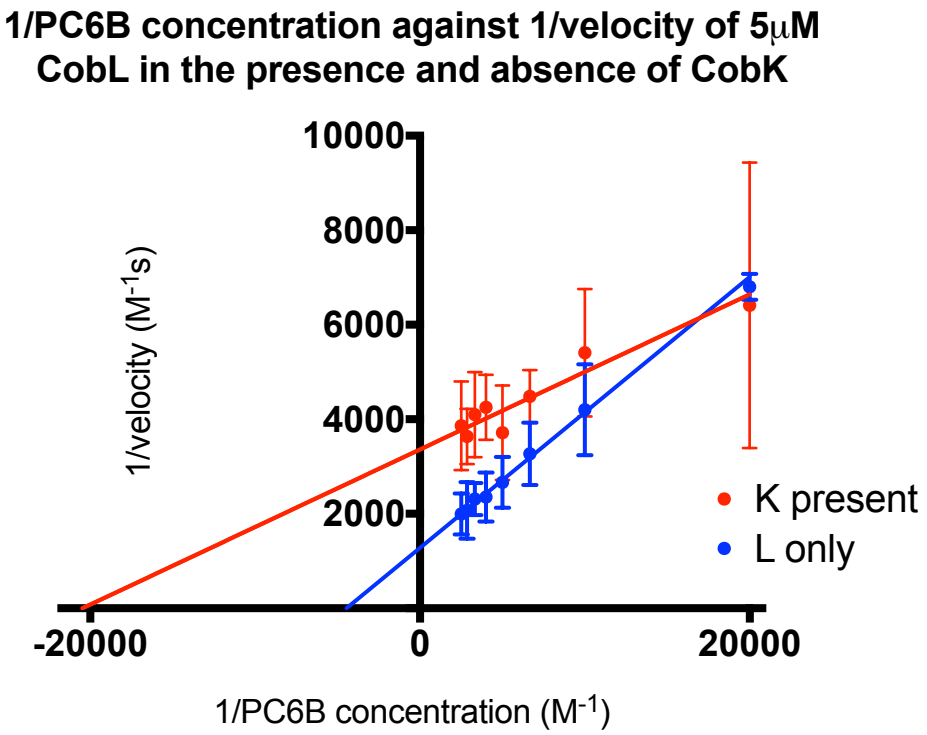


Figure 66: A) Shows a Michaelis-Menten kinetics representation of PC6B turnover by CobL in the presence (red) and absence (blue) of CobK, where initial velocity was plotted against substrate concentration. B) Shows a Lineweaver-Burk plot of PC6B turnover by CobL in the presence (red) and absence (blue) of CobK.

	CobL only	CobK present
V_{max} (Ms^{-1})	784 x 10^{-6}	298 x 10^{-6}
K_M (M^{-1})	225 x 10^{-6}	48.9 x 10^{-6}
k_{Cat} (s^{-1})	15.7	6.0

Table 10: A summary of the V_{max}, K_M and k_{cat} for PC6B turnover in the presence and absence of CobK where V_{max} is the maximum enzyme velocity.

4.3.7 NMR of CobK with CobL^{c}

1D proton NMR spectra recorded from 0.05 mM CobK with a ligand concentration of 0.15mM NADPH and PC6B is shown in Figure 67. After this solution had been scanned a 0.15mM excess of CobL^{c} was added to this solution and a spectrum recorded. Finally, an excess of 1mM SAH was added to this solution to mimic the addition of SAM and arrest PC6B turnover during the experiment. Upon addition of SAH three new peaks can be observed between 8.5-9.2ppm which are in the region for the corrin ring of PC6B, which indicates that in the presence of CobL^{c} SAH is released. Full length CobL was not used for this experiment as the tetramer is too large.

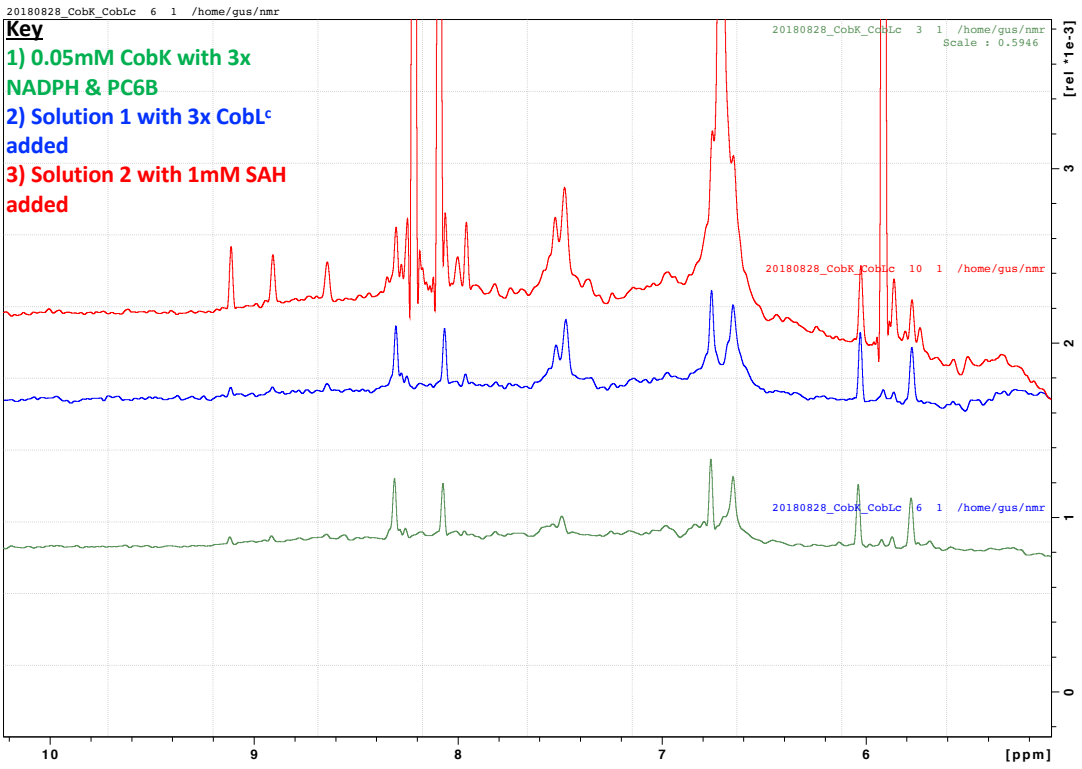


Figure 67: 1D NMR of CobK with PC6B and NADPH bound (green) with a stepwise addition of CobL^{c} (blue), followed by SAH (red). Upon the addition of SAH three peaks can be observed 8.5-9.2ppm indicating the release of PC6B.

4.4 Promiscuous binding of cobalamin intermediates to CobK

4.4.1 Fluorescence studies

When HBA is placed under a UV lamp it shows fluorescent activity, while PC6B does not. However, when HBA is in the presence of CobK the fluorescence changed from orange to green, Figure 68. This suggests that HBA binds to CobK. To explore this a UV-VIS wavelength scan was used to observe if there was a wavelength shift similar to when PC6B binds to CobK. This result is shown in Figure 68 and shows there is no observable absorption change between free HBA and HBA in the presence of CobK.

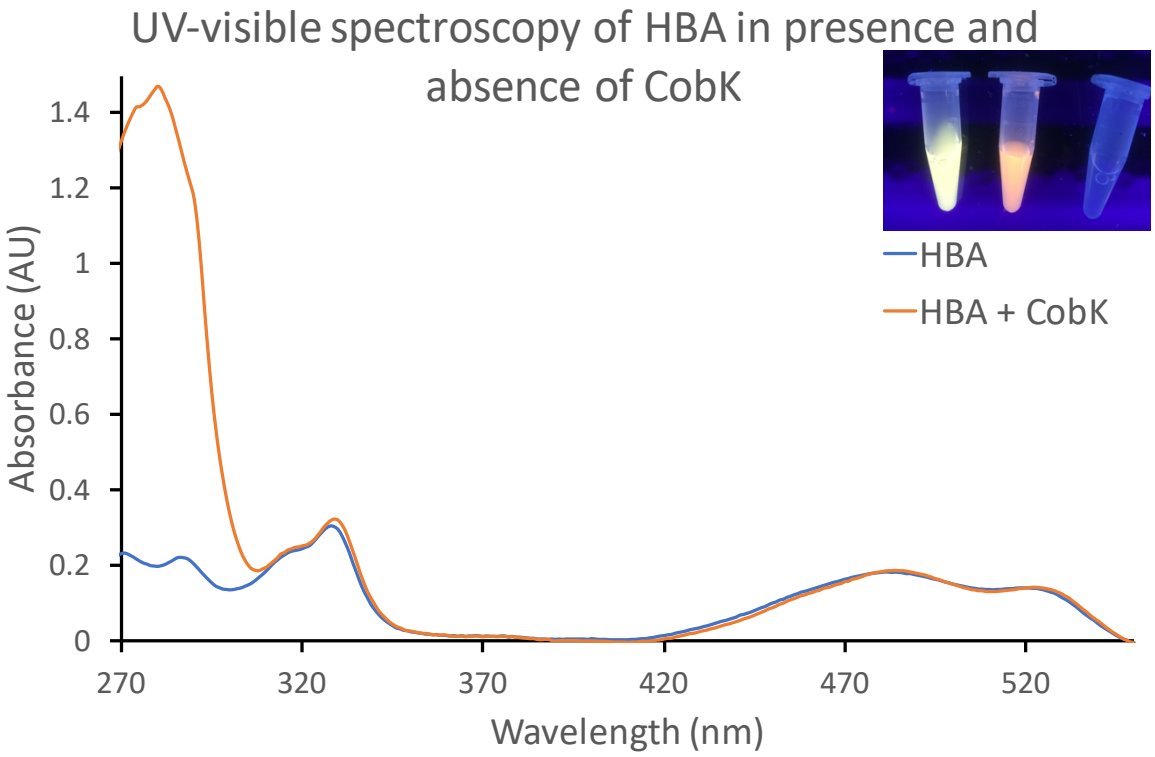


Figure 68: Wavelength scan of HBA in the presence and absence of CobK to see if there is a change in absorbance (as observed with PC6B). It was noticed that under UV light the fluorescence of HBA changes from orange (middle tube) to green (left tube) in the presence of the protein CobK. The right tube is CobK protein only, showing fluorescence is not due to protein.

To further test the potential binding of HBA to CobK a fluorescent spectroscopy experiment using nickel beads was undertaken to show if CobK was in fact able to bind HBA. The experiment used three separate samples of CobK, HBA and CobK+HBA. These were mixed into separate Eppendorf tubes containing 200µl of charged nickel beads and CobK loading buffer (buffer A). The solutions were mixed and then spun at a speed of 1000rpm for two minutes to pellet the beads. The tubes were photographed under a UV lamp, before the supernatant was collected for fluorescence spectroscopy and a sample taken for an SDS-PAGE gel. The process was repeated by adding increasing imidazole concentrations using buffer B and buffer C as described in section 2.4.1. Two excitation wavelengths were used when scanning HBA samples using fluorescence spectroscopy as the spectrum profile contains two absorption peaks at wavelengths 475nm and 525nm.

The results are shown in Figure 69. The SDS-PAGE gel confirms CobK bound to the nickel beads and was not eluted into the supernatant until the final step. The fluorescent lamp photos show that there was excess CobK protein with respect to nickel beads as observed in the fluorescent flow through supernatant and SDS-PAGE gel. The most intriguing image was the wash step where the pelleted nickel beads fluoresce at the bottom of the tube, showing CobK and HBA are bound, while for the HBA sample the entire tube is fluorescent showing HBA does not bind to the nickel beads.

However, the fluorescent spectroscopy spectra do confirm the presence of HBA in the elution step when mixed with CobK via a higher excitation signal than the wash step at both wavelengths. Combined with the results in Figure 69A, this confirms that CobK retains HBA while binding to the nickel beads. The fluorescent spectroscopy profiles of HBA flow through, wash and elution show a decrease in the HBA signal at each step, confirming that HBA is not able to chelate to the nickel beads. It is also interesting to compare the fluorescent profile of HBA when excited at 475nm, as the peak observed at 600nm disappears in the presence of CobK further supporting it is able to bind to CobK, as shown in Figure 69Bi.

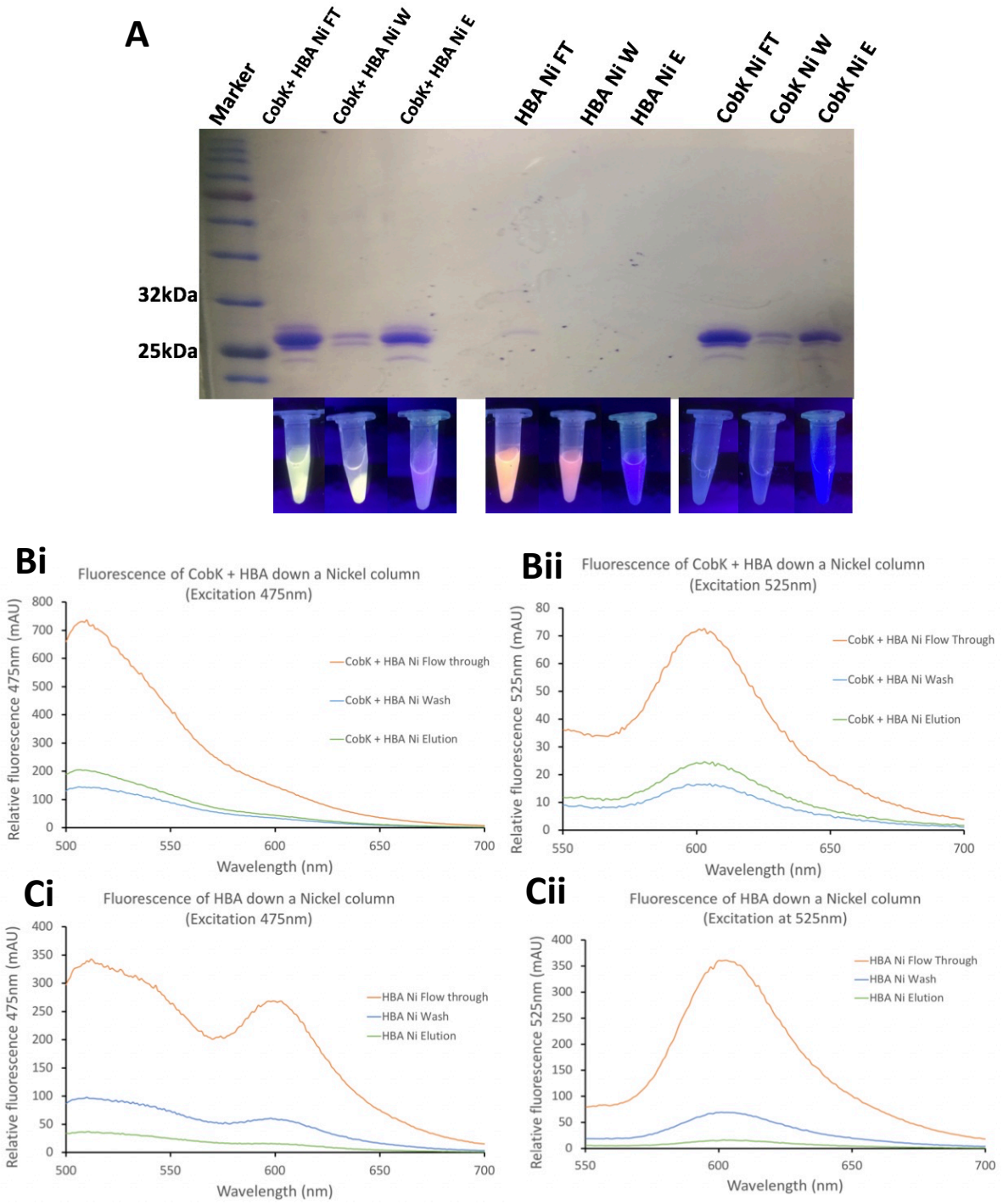


Figure 69: CobK binding HBA nickel pull down. A) Shows an SDS-PAGE gel of the flow through, wash and elution steps of the nickel pull down for samples CobK+HBA, free HBA and free CobK accompanied with fluorescence observed under a UV lamp. B) Shows a fluorescence wavelength scan of CobK+ HBA nickel pull down excited at 475nm (i) and 525nm (ii). C) Shows a fluorescence wavelength scan of HBA nickel pull down excited at 475nm (i) and 525nm (ii).

4.4.2 NMR of CobK binding HBA

A 2D NMR HSQC experiment was carried out using N^{15} labelled CobK to observe if the protein was able to bind HBA. Three samples were measured: native CobK, CobK + NADPH + PC6B and CobK + NADPH + HBA. The spectra are shown in Figure 70 and all show different conformations of CobK, as they do not fully overlap. When the CobK+HBA spectrum is overlaid with the NMR spectrum from section 4.2.5 of CobK binding PC6B a very similar spectrum is seen suggesting HBA has bound to CobK. This observable peak shifting supports binding of HBA to CobK.

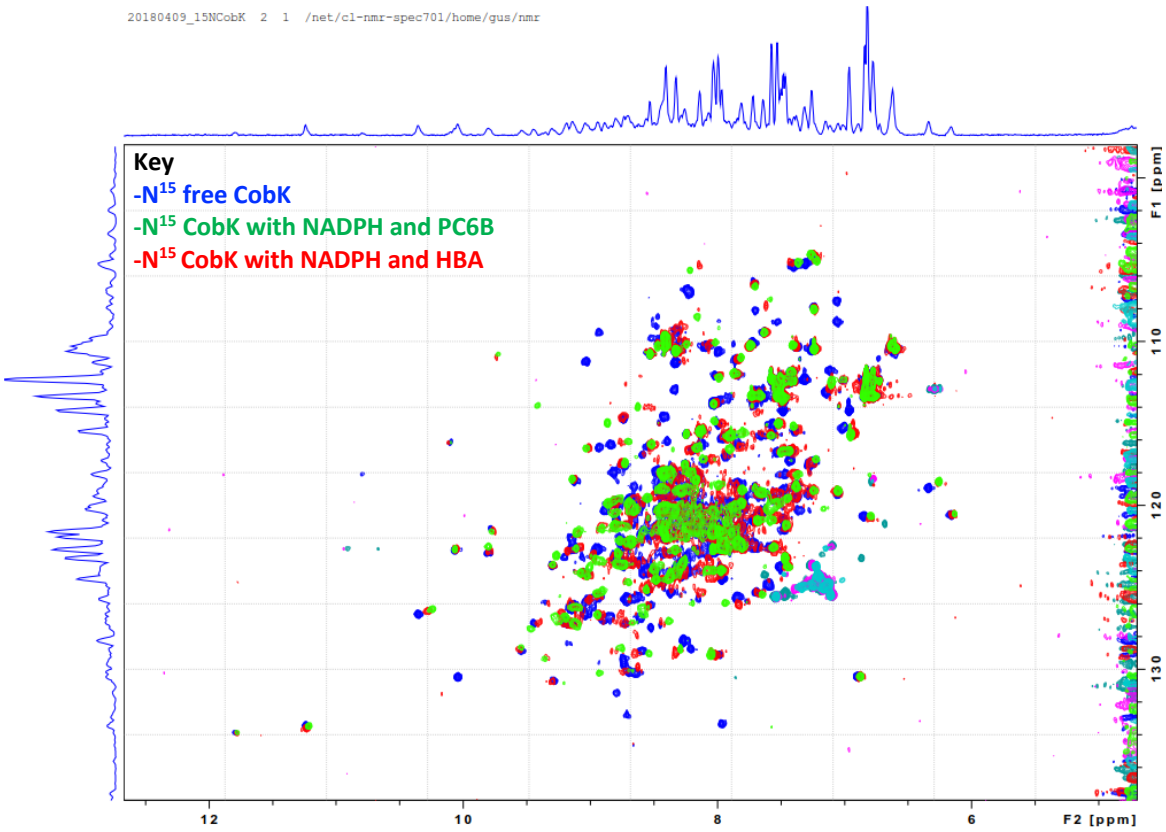


Figure 70: HSQC NMR of N^{15} labelled CobK in the presence of ligands NADPH+ PC6B or NADPH and HBA. There is a clear change in spectra between native CobK protein (blue) and ligand bound protein (red/green). The spectrum of PC6B (green) and HBA (red) samples are remarkably similar showing CobK is able to bind HBA.

4.4.3 CobK crystals with ligand binding

Finally, crystallography was used to confirm visually HBA binding to CobK. CobK protein was purified and crystal trays set with both the ligands HBA and PC6B, as the current published structure contains C5-methylated PC6B. HBA binding to CobK was inferred during the protein concentration step, as the pink pigment remained in the upper half of the concentrator. Pink crystals did not form when CobK was mixed with HBA. However large crystals of CobK ranging between 1-4mm in length in the presence of 5x excess ligand concentration of PC6B and NADPH were obtained, Figure 71. The crystal conditions used were 17-19% w/v PEG 3350 and 0.2M NH_4Cl pH8.0, with a protein concentration of 20mg/ml. These crystals were picked in a cryo-solution containing 20% glycerol and sent off for data collection at Diamond Synchrotron, where the data collection and refinement is summarised in Table 11.

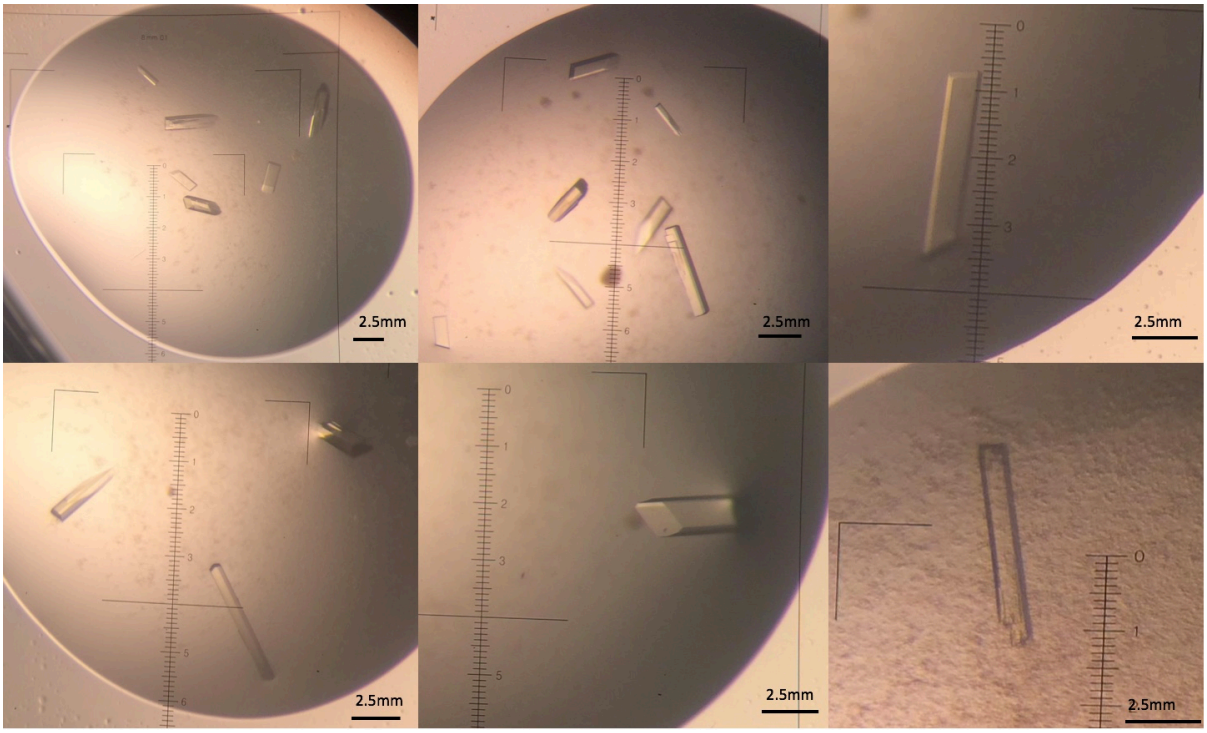


Figure 71: CobK crystal optimisation produced in the presence of PC6B and NADPH

CobK+ PC6B data collection statistics

Space group	$P\ 2_1\ 2_1\ 2_1$		
Cell edges (Å)	33.857	73.885	103.612
	90.000	93.293	90.000
Wavelength (Å)	0.97925		
	Overall	**Inner shell**	**Outer shell**
Resolution (Å)	1.16-60.16	3.15-60.16	1.16-1.18
Number of observations	486,306	25,482	20,347
Number of unique reflections	91,042	4,837	4,506
Multiplicity	5.3	5.3	4.5
Rmerge	0.074	0.034	1.772
Rpim	0.035	0.016	0.928
Rmeas	0.082	0.037	2.007
Mean I/σ	9.1	38.1	0.6
$CC_{1/2}$	0.999	0.998	0.328
Completeness (%)	99.9	99.0	99.8
Wilson B estimate ($Å^2$)	11.0		

Refinement statistics

Resolution (Å)	1.17-60.16
Number of reflections	88,385
R-work/R-free	0.20/0.22
Number of atoms (protein chains/ligand/water)	1/2/441
rmsd bond length	0.010
Rmsd bond angle	1.343
Mol-probability score	1.19
Residues in most favoured regions (%)	99.19
Outliers (%)	1.13

Table 11: Crystal data and refinement statistics for CobK with PC6B and NADPH. Rmerge = Σ hkl Σ i |Ii–< I>| /– Σ hkl Σ Ii, where Ii is the intensity of the ith observation, <I> is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively and Rpim is a measure of the quality of the data after averaging the multiple measurements and Rpim = Σ hkl [n/(n-1)]1/2 Σ i |Ii(hkl)–< I(hkl) > |/Σ hkl Σ i Ii(hkl), where n is the multiplicity, other variables as defined for Rmerge R-factor = Σ hkl |Fo–Fc |/Σ hkl Fo, where Fo and Fc represent the observed and calculated structure factors, respectively. The R-Factor is calculated using 95% of the data included in refinement and R-free the 5% excluded[62,103].

Upon solving the structure of CobK with PC6B it was in fact confirmed to be methylated at the C5 position, Figure 72A in a similar fashion to the already published structure of product bound CobK. Given the crystallisation conditions used were identical to the published structure, the unit cell dimensions and space group were identical. When these two CobK structures were compared there was little difference between them, with only a small degree of variation within some of the loop regions, Figure 72B, with both of these structures having been solved to a similar resolution of ~1.2Å.

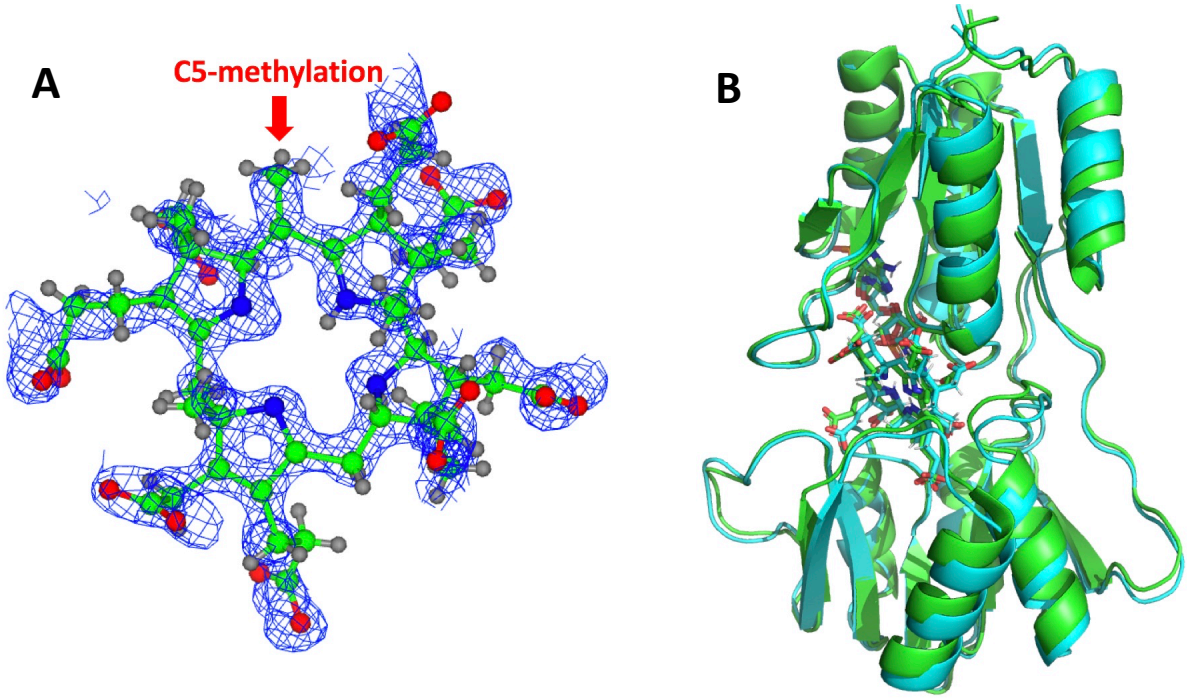


Figure 72: A) Electron density map of C5-methylated PC6B clearly showing the presence of a methyl-group at the C5 position of PC6B. B) Structural comparison of the product bound CobK structure solved during this project (blue) overlaid with the existing published CobK PDB structure 4X7G (green), showing a degree of flexibility within some of the loop regions. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.This alignment gave an RMSD of 0.880 when 240 of 249 alpha carbons were aligned using the Pymol function 'align Prot1////CA, Prot2, object=alignment'.

4.5 Discussion and future work

4.5.1 CobK product binding activity

It is clear that CobK is able to hold on to its product PC6B and will in fact re-bind it in the presence of NADPH. PC6B is structurally very similar to PC6A only differing by a single hydrogen on carbon-19. So, it is questioned whether CobK product binding activity is important for some form of protection. PC6A is an unstable intermediate and therefore not routinely purified in the laboratory whereas PC6B appears much more stable. Perhaps tight binding is to ensure that PC6A is protected until NADPH becomes available or it facilitates the safe transfer of product to CobL, ensuring that the product is not lost or exposed to off-pathway reactions.

Product binding was demonstrated using a range of techniques including UV-VIS spectroscopy, size exclusion chromatography, thermal melts, 2D NMR and ITC. Through UV-VIS spectroscopy the absorption maximum of PC6B was shown to change in the presence and absence of CobK protein. This was demonstrated through a titration where a clear shift between free PC6B to bound PC6B was observed as CobK was added. An isosbestic point was seen indicating that two states of PC6B exist in the solution; PC6B bound to CobK and PC6B free in solution. This titration also demonstrated binding of PC6B to CobK was more likely in the presence of co-factor NADPH, as a repeated titration carried out under identical conditions with the removal of NADPH resulted in a lower absorption maximum for bound PC6B to CobK at 348nm.

PC6B product binding was further supported through thermal melt assays. These showed that the CobK will bind both ligands separately, while the presence of both ligands resulted in the largest increase in protein stability, where the T_m of native CobK increased from 37^{o}C to a T_m of 53^{o}C. It was interesting to note that native CobK had a T_m of 37^{0}C as this meant that the protocol for preparing N^{15} CobK for NMR experiments was altered as the protocol suggested growing cells at 37^{o}C, so this was lowered to 18^{o}C [99]. This observation could also explain why Fang et al, 2018, struggled to

produce *de novo* biosynthesis of cobalamin in *E.coli* when grown at 37^0C as they would have denatured at least half of CobK present, greatly hindering the cobalamin production pathway[113].

Size exclusion chromatography also gave a longer retention time of CobK when ligands were present suggesting that the protein becomes tighter and therefore more rigid as the flexible loop region closes over the active site, pulling the two Rossman domains together. A change in conformation can be observed when the crystal structures of native and ligand-bound CobK are overlapped. However, the SEC highlights that the native CobK may in fact be quite flexible giving a more dynamic picture of the protein that X-ray crystallography cannot provide.

The incorporation and production of N^{15} labelled CobK was successful and allowed 2D NMR HSQC spectra to be recorded. This demonstrated that CobK was able to bind NADPH and PC6B, both of which caused a conformation change indicated via a change in the shift values of the spectra, with the most dramatic shift observed when both ligands were present. Despite amino acid residues not being assigned it was still apparent that a change in conformation is occurring, which supports the flexible loop observed in the ligand bound crystal structure, but which is not present in the native crystal structure.

In an attempt to quantify product binding activity and determine a dissociation constant (K_d), ITC was used. It was interesting to observe that PC6B binding was an endothermic process, with binding being driven by the release of water molecules from the active site, suggesting this is being driven by entropic factors rather than enthalpy. A K_d constant of 3μM was obtained. Despite the similarity between PC6A and PC6B it would be expected that CobK has a higher affinity for its substrate than product. However, it should also be noted that the stoichiometric ratio where n= 0.85 was obtained. This could mean that the product does not fully bind to CobK, however it is more likely that this is due to error in determining the concentration of PC6B stocks, as weighing out the powder would have introduced a large degree of error. Any error in the stoichiometric ratio *n* will consequently result in greater error in the derived thermodynamic parameter calculated during ITC[114]. Therefore, it does not instil confidence in the K_d value quoted. However, this experiment still demonstrates that

CobK exhibits product binding activity, supporting the other techniques used, but further experiments could be undertaken to improve these thermodynamic measurements. This could include measuring the K_d of each ligand separately, but also extended to probing a potential protein-protein interaction between CobK and CobL, discussed below.

4.5.2 Probing potential protein-protein interactions between CobK and CobL

Initial techniques to probe a protein-protein interaction between CobK and CobL included using native-PAGE gels, a nickel-column pull down and size exclusion chromatography experiments, all of which were unable to show that CobK and CobL interact. In these experiments when ligands were used NADPH was added to encourage PC6B product binding to CobK and SAH was used in place of SAM to prevent the production of PC8, in the hope that it would assist in trapping the proteins together rather than PC6B turnover. In the Native-PAGE experiment an equimolar concentration of CobK:CobL was trialled, but given CobL is a larger protein and forms a tetramer it is possible that four CobK enzymes may be able to bind at once around the region of the first active site located on the C-terminus of each monomer, hence why increasing CobK concentrations were explored. The nickel-column pull down again did not remotely suggest an interaction may be present, but it did demonstrate that the Histidine tag can be successfully removed from both CobK and CobL using thrombin at an optimum temperature of 16^0C overnight. The SEC studies did not show a protein-protein complex via an earlier elution profile but did surprisingly end up highlighting the dynamic nature of CobK upon PC6B product binding which was discussed above in section 4.6.1. When a thermal melt study was undertaken on CobK and CobL together with the ligands NADPH and PC6B, to encourage product binding, and SAH or SAM, different melting temperature (T_m) values were obtained that were not in agreement when individual protein and ligands were used. When only one ligand was used in combination with CobK and CobL together (data not shown) the T_m remained at

44°C, therefore possibly indicating that a protein-protein complex is being formed with a different degree of stability to the individual proteins.

Having researched around CobB and the action of class I amidotransferases, discussed in chapter 1 section 1.4.5, and how the two active sites wait for the other to be occupied before carrying out catalysis, it was thought a similar action may occur via a CobK and CobL complex. There is a small structural similarity between the ligands NADPH and SAM/SAH, which both share an adenine joined to a ribose unit. Therefore, PC6B may not be released from CobK until SAM is present for CobL activity. This was investigated by measuring the shift in the absorption maximum of bound PC6B at 348nm and the expected absorption maximum of free PC6B at 378nm. As increasing concentrations of SAM and SAH were titrated into a solution containing bound PC6B to CobK, an absorption shift was never observed and does not trigger the release of PC6B. Given how buried the binding site of SAM is in the crystal structures of methyltransferases, shown in **Figure 10** and Figure 15 in Chapter 1, this result is not too surprising.

When PC6B is converted into PC8 via CobL there is a shift in the absorption from 348nm to 390nm. This was used to measure the rate of conversion for CobL. As PC6B has a higher absorption maximum the wavelength 348nm was therefore selected to follow the reaction over time. This conversion was carried out in the presence and absence of CobK to see whether a change in rate was observed and indicate if a potential protein-protein interaction is occurring.

Pre- and post-scans confirmed the conversion and whether PC6B was bound to CobK if present. The absorption maximum of PC6B in CobL buffer of pH8 did shift to 348nm, Figure 73, but was not as pronounced as CobK buffer of pH 7.5. Finally, a large excess of SAM was added to initiate the experiment and also provide an almost consistent concentration so that it could be ignored for the kinetics analysis.

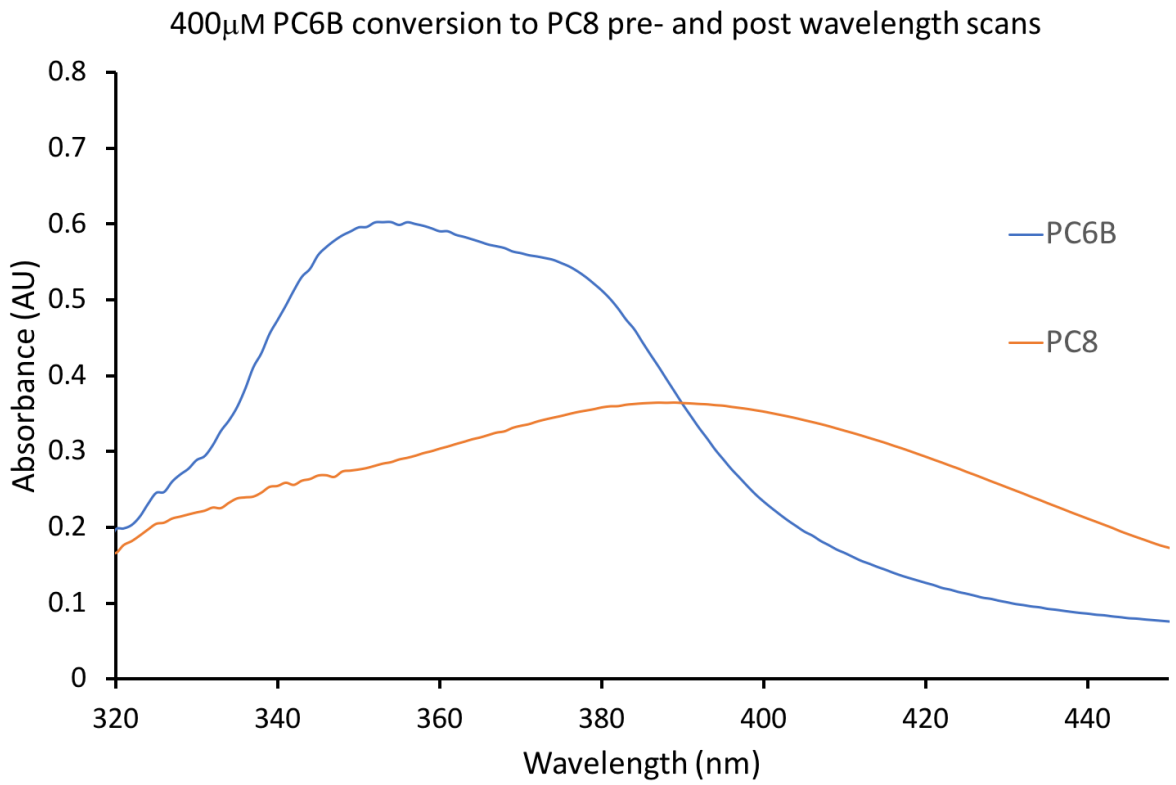


Figure 73: PC6B pre-scan and PC8 post-scan of PC6B conversion with CobL only, in CobL buffer of pH8.0

The kinetics analysis of PC6B conversion shown in Figure 66 and Table 10 does appear to suggest that CobK and CobL do interact. If CobK binds to CobL this would reduce the concentration of active CobL to react with substrate which causes a reduction in V_{max}. When CobK is present the reaction reaches a half V_{max} at a lower substrate concentration which changes the k_m. This may mean that the presence of CobK affects the affinity of CobL for its substrate PC6B. CobK may be binding to CobL in such a way that it inactivates the active centre, trapping the enzyme-substrate complex, indicating that CobK is acting in a similar fashion to an uncompetitive inhibitor, where both the V_{max} and K_m values are changed[115].

At high CobK-PC6B concentrations the V_{max} falls by a factor of approximately two-fold and the K_m falls by roughly a factor of four. If K_m were unchanged there would be no effect on substrate binding, however it is interesting that binding appears to be tighter. The k_{cat} for the reaction does not dramatically change. Therefore, the results suggest the possibility of higher affinity for substrate and

a reduction of the number of active sites available which would be compatible with the direct transfer of PC6B from CobK to CobL.

The 1D proton NMR experiment also supports a CobK-CobL interaction. Protein concentrations were at the lower limits of the technique but when SAH was added to a mixture of CobK with PC6B bound and $CobL^c$ a change in the proton spectra was observed where three new peaks were observed between 8.5-9.2 indicating the release of PC6B into solution.

The kinetics and NMR experiments are both in the early stages of concluding a protein-protein interaction but are certainly promising. Improvements for further work would include testing the kinetics on $CobL^c$ as this is the active site that acts first on CobL. However, the absorption spectrum of PC7 would need to be measured so that a suitable wavelength can be selected during the kinetic measurements. Also, as the CobL concentration used during the kinetics was quite low it is possible that using the same buffer in the presence and absence of CobK at pH7.5 may not cause it to precipitate out of solution and would therefore provide a much better comparison between the two experiments regarding the initial PC6B absorption spectrum. This may also explain why the residuals plot observed when fitting initial absorbance against time for CobL became worse at higher concentrations, as the same fitting equation was used for both experiments.

Structural insights into a possible CobK-CobL complex would allow further study via targeted mutagenesis, but co-crystallisation attempts were unsuccessful. Initial NMR studies appear to support an interaction but are limited to smaller protein sizes. This limits the technique to CobK-$CobL^c$ studies, which could certainly be repeated at higher protein concentrations. If amino acid residues were assigned to a HSQC experiment using N^{15} or C^{13} labelled CobK this would allow the identification of interacting amino acids to be identified. An interaction between these enzymes would most likely be acting on opening the flexible post-β2 loop region of CobK. The orientation of PC6B, when bound to CobK, is such that the C15 methylation site is facing outwardly and therefore it is plausible to speculate that it is presented to CobL. For example, this could be instigated via a

competing arginine on CobL, which out-competes an existing arginine interaction on CobK which can be seen acting as a latch for loop closure, as highlighted in Figure 11 in Chapter 1.

4.5.3 Promiscuous binding of cobalamin intermediates to CobK

HBA is a fluorescent orange/pink cobalamin intermediate and it was observed that the fluorescence changes to green in the presence of CobK protein. A UV-VIS spectroscopy wavelength scan was initially measured on a sample of HBA with and without CobK present to see if there was a change in absorbance, similar to when PC6B is free in solution or bound to CobK. However, spectra of HBA did not change when in the presence of CobK. To investigate this further a nickel column pull down experiment was carried out to see if when HBA and CobK were mixed, HBA would be present when CobK was purified via its histidine tag. A control containing only HBA was carried out in tandem to confirm that HBA was not able to chelate nickel, where it would be purified alongside CobK rather than through CobK binding. This was tracked using an SDS-PAGE gel and visually checking the fluorescence under a UV light. CobK acted as expected during the purification and it did appear that CobK was binding HBA, which is best observed in Figure 69 during the wash step and the fluorescence is visible on the pelleted nickel beads, which did not occur in the HBA wash step.

The supernatants from each step during the nickel pull down experiments were collected and measured using fluorescence spectroscopy. As HBA has two absorption maximums at 475nm and 525nm these were both used separately as excitation wavelengths. When HBA was excited at 475nm in the presence of CobK, an absorption emission maximum around 600nm was not present when compared to the HBA control. At both excitation wavelengths the nickel column elution measurements contained more HBA than the wash step when CobK was present, where excitation at 475nm gave a stronger signal.

These two experiments do appear to support the idea that CobK is incorrectly binding HBA. However, it was found that as the imidazole concentration increased throughout the pull down it did appear to interfere with the visual fluorescent profile of HBA, Figure 69 elution profile, which also

prevented successful UV-VIS wavelength scans. The volumes in this experiment were very small, but if repeated the imidazole should be removed by buffer exchange or a desalting column to provide better spectroscopy profiles.

This was followed up with structural studies using 2D NMR in a similar fashion to section 4.3.5. N^{15} labelled CobK was produced and spectra profiles of native CobK, CobK with ligands NADPH and PC6B and CobK with ligands NADPH and HBA were compared. The spectra of native CobK varied greatly from the other two which were overall highly similar. This is in agreement with the conformation change observed when the flexible loop region of CobK binds ligands and further supports incorrect binding of HBA.

Production of CobK crystals with HBA were also trialled to explicitly confirm binding but failed. Native CobK crystals were also attempted in the hope of soaking with HBA, but again failed to produce anything. In tandem with these trials CobK crystals with the ligand PC6B were also attempted as the current structure solved by the Pickersgill group in fact contains C5-methylated PC6B[62]. This did produce crystals, where further attempts to soak with HBA were also trialled but were also unsuccessful.

The protein structure solved that was put down for crystallography trials was purified directly from PC6B producing cells, rather than adding purified PC6B, which has been confirmed by mass spectroscopy[53]. This plasmid contains all the Cob enzymes up to CobK, so multiple methyltransferase enzymes are present and may be able to act out of turn and incorrectly methylate C5. Given PC6B does not share the same molecular mass with preliminary cobalamin intermediates in the pathway, this rules out any out of turn methylation occurring during the prolonged 24hr cell growth, but rather after CobK purification. The protein purification was followed using SDS-PAGE gels so it is possible a methyltransferase was purified alongside CobK as their molecular masses only differ by ~1kDa, which may not differentiate on the resolution of a gel. However, none of the Cob enzymes before CobK contain a histidine tag and the canonical methyltransferases form dimers while CobK is

a monomer. This would have led to two elution profiles which were not obtained during size exclusion chromatography making this C5 methylation difficult to explain.

Regardless the C5-methyl group is clearly shown in the electron density map of the crystal structures highlighting that CobK does not only bind its substrate PC6A and its product PC6B but is also able to bind two other corrin ring based intermediates; HBA and C5-methylated PC6B. As to which methyltransferase has acted out of turn, CobM is initially proposed based on comparison of comparing the binding ligands. CobK binds a contracted corrin ring of nineteen carbons narrowing the choice to CobF and CobM. CobF acts on removing the ethanone decoration on C1 before carrying out a methylation. Further methylation has been identified before by Warren et al, 1990 where CysG (CobA) was found to bring about a third methylation at a high concentration to C12 of PC2[55]. The quickest method to confirm this would be to purify all the methyltransferases separately and add PC6B followed by mass spectroscopy to confirm whether an additional methylation has occurred.

4.5.4 Concluding remarks

A wealth of experimental techniques have been undertaken and reported in this chapter to predominantly explore CobK binding interactions. The unusual product binding activity of PC6B to CobK has been demonstrated. In an attempt to answer how CobL obtains its substrate PC6B from CobK, a potential protein-protein interaction has been explored between these two enzymes. Attempts to trap the complex structurally via X-ray crystallography were unsuccessful but NMR and kinetics do provide supporting evidence of an interaction. Further kinetic studies and ITC would continue to probe this. The lack of a full length CobL structure makes targeted mutagenesis difficult to track down key binding residues, but the CobL^{c} structure could assist in initial attempts.

To conclude preliminary studies support HBA binding to CobK, which is produced further upstream in the biosynthesis of cobalamin. This chapter has demonstrated that out of turn HBA binding and an additional methylation can occur within the reaction pathway. These observations

coupled with the fact that the substrate PC6A is an unstable compound, add weight to the idea that product trapping activity of CobK could be important for substrate channelling and protection of out of turn reactions within the cobalamin pathway itself. Therefore, CobH may also prevent other out of turn modifications to HBA and ensure the safe delivery of HBA to CobB. This has led on to the investigation of how specificity of cobalamin intermediates is brought about and is discussed in the next chapter, with a particular focus on the canonical methyltransferases.

Chapter 5: The Cob Methyltransferases

5.1 Introduction, aims and objectives

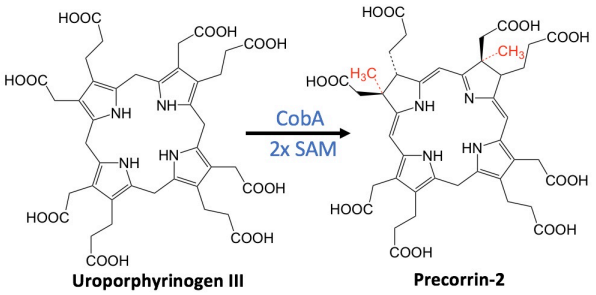
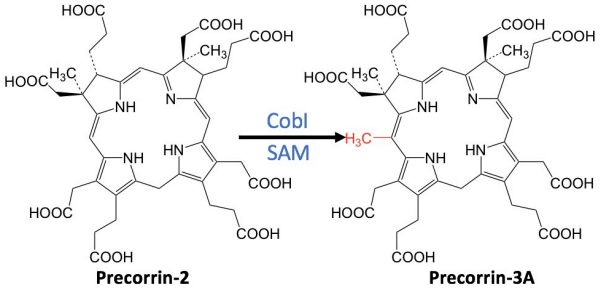
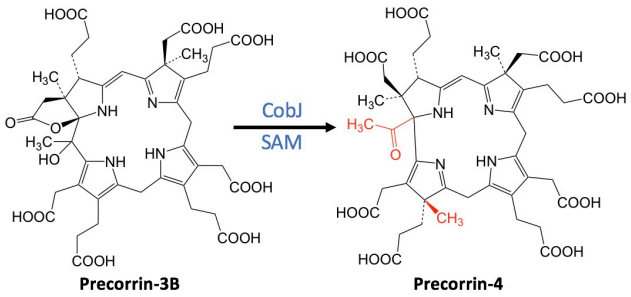
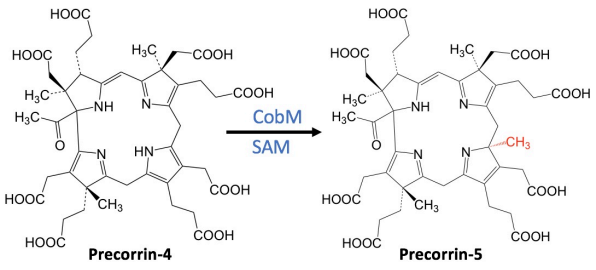
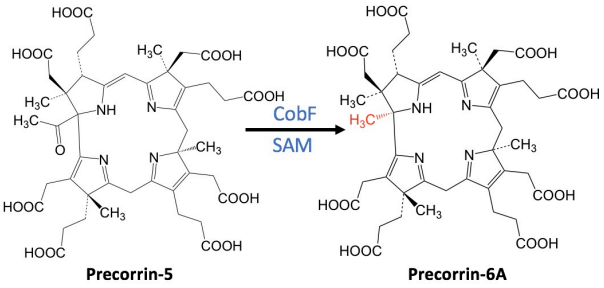
Methylation is a key reaction required in the development of the cobalamin corrin ring, requiring eight methylations in the transformation of uroporphyrinogen III to hydrogenobyrinic acid[1,15]. These are catalysed by the six so called canonical methyltransferases and CobL^{c} a non-canonical methyltransferase, as summarised in Table 12. The canonical methyltransferases are classified via their tertiary fold, in which the active site is located between two $\alpha\beta\alpha$ domains each containing five β-strands[116].

These methylation steps are carried out via the transfer of a methyl-group donor S-adenosyl-methionine (SAM). It has been discussed in section 4.3.3 that at least two of the cobalamin biosynthetic enzymes are prone to mistaking some of the cobalamin intermediates. Therefore, to conclude this project it was of interest to focus on exploring how the cobalamin methyltransferases have developed their regio- and stereo-specificity. To explore specificity regarding precorrin cobalamin intermediates of the canonical methyltransferases the following areas were studied:

1. A sequence analysis to compare the anaerobic and aerobic pairs leading to a discussion of methyltransferase structure studies.
2. Those methyltransferases which have had their structures solved were further explored using docking software and outputs which had already been generated by the Pickersgill group but not analysed in depth. This unpublished work contained information on different docked precorrin ligands. The analysis involved identifying the most feasible models which were further studied through identifying and summarising binding interactions and the corrin ring decoration. It is proposed that the decoration groups of

the corrin ring of these precorrin ligands may help drive specificity as demonstrated in Figure 74.

3. Having identified loop regions involved in precorrin docking across the canonical methyltransferases, their sequences were further explored in an attempt to conclude whether conserved residues are important for substrate specificity.

Cob enzymes	Addition of methyl-group to	Ancillary reaction	Graphic
CobA	C2 & C7	-	
CobI	C20	-	
CobJ	C17	Ring contraction	
CobM	C11	-	
CobF	C1	C1 acetic acid extrusion	

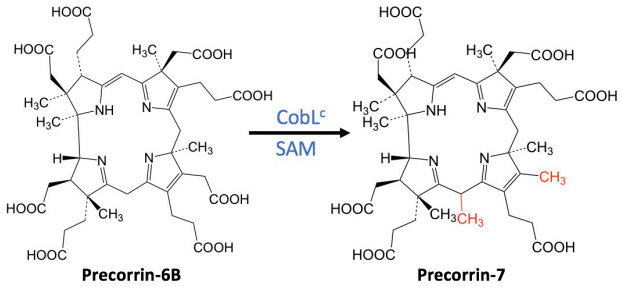
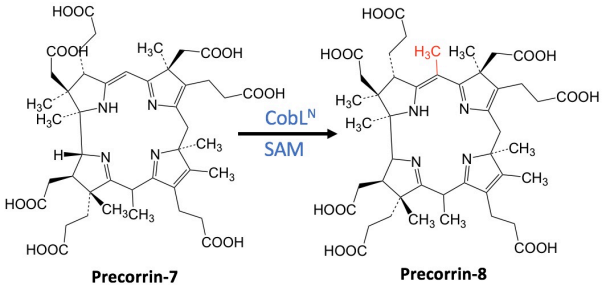
CobL^{C}	C15	C12 de-carboxylation	 Precorrin-6B Precorrin-7
CobL^{N}	C5	-	 Precorrin-7 Precorrin-8

Table 12: A summary of the aerobic methyltransferases and the reaction(s) they catalyse, which are shown in red.

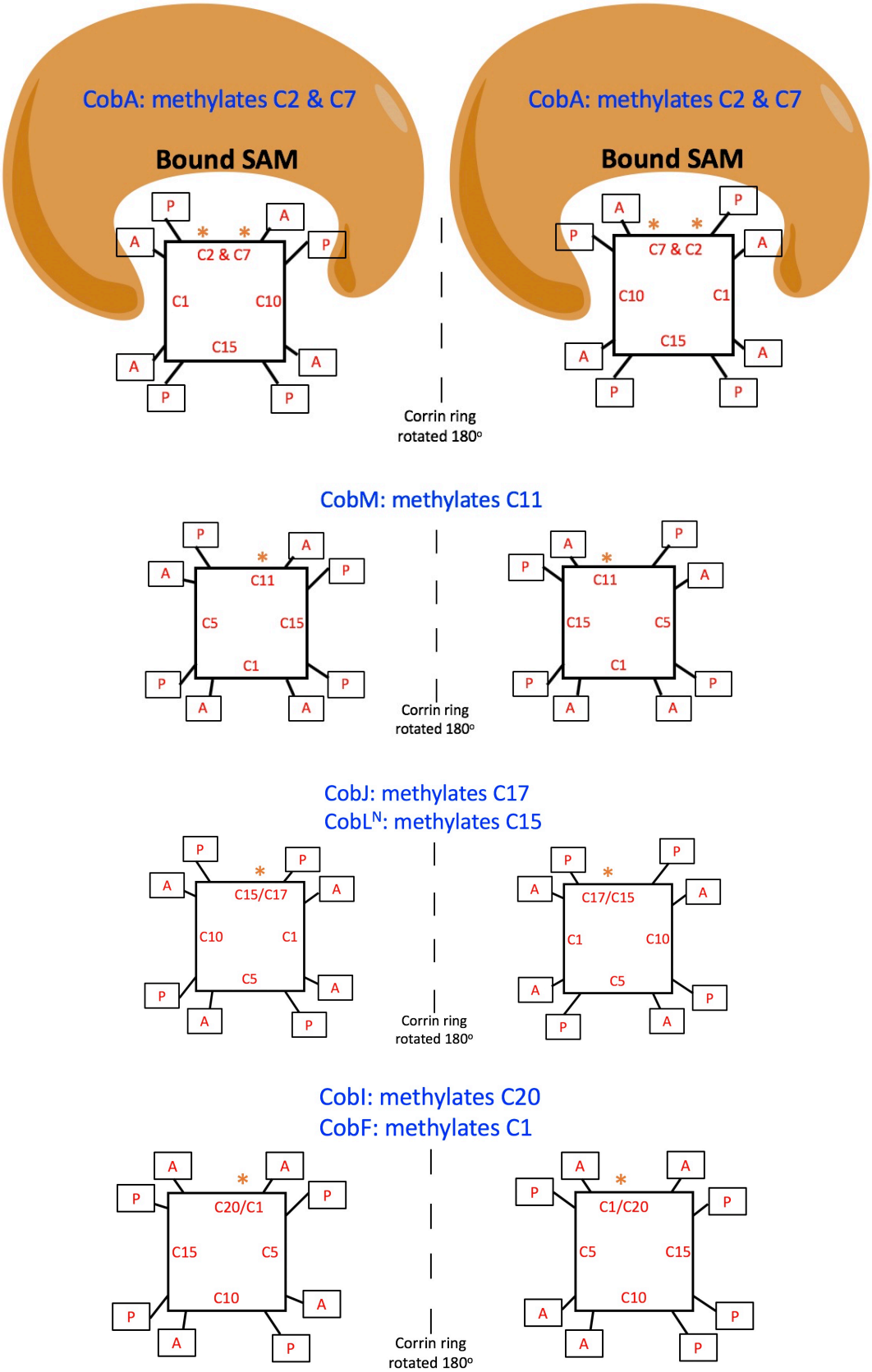


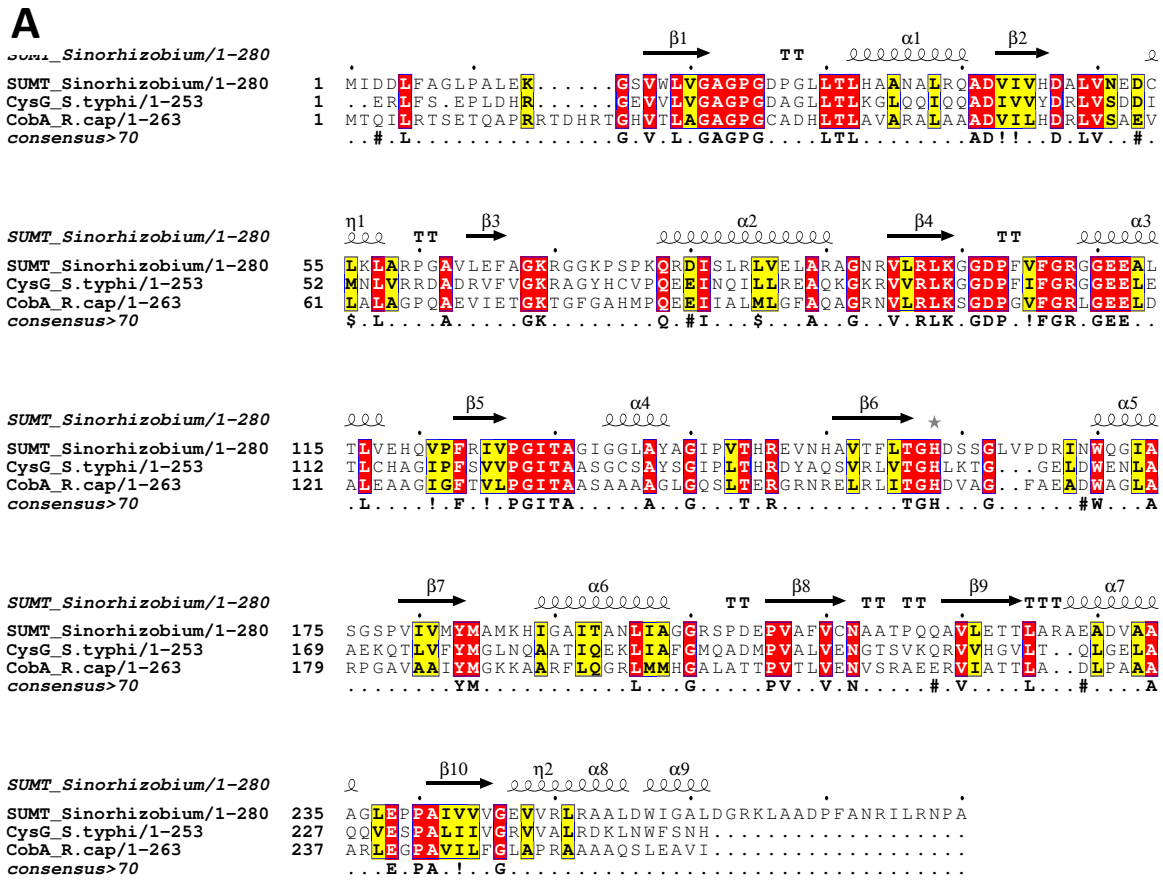
Figure 74: A simplified precorrin ligand docking schematic. Methylation is assumed to occur on the North edge demonstrating how 180^0 rotation in the y-axis would change the decoration of the corrin ring.

5.2 Sequence analysis of aerobic and anaerobic methyltransferase pairs

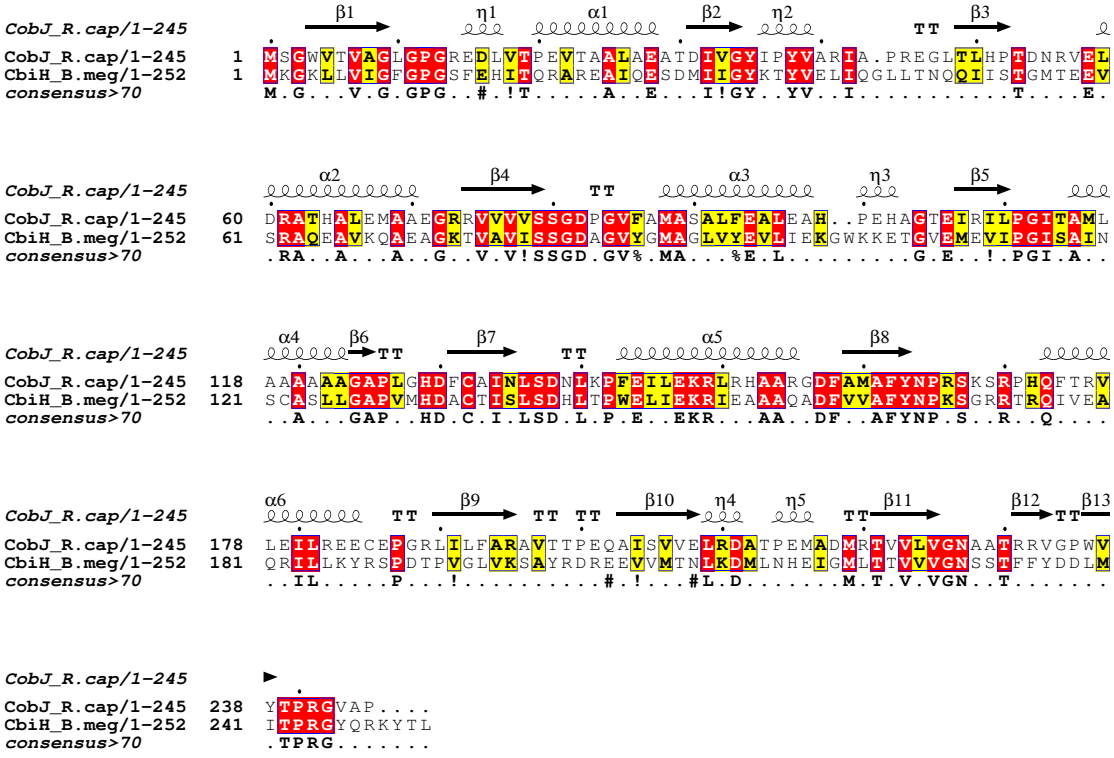
In the first part of this analysis a sequence comparison of the canonical methyltransferases was undertaken to observe if any conserved amino acid residues could be identified for binding precorrin intermediates. This was initially achieved by comparing the aerobic pathway, using gene sequences from *Rhodobacter capsulatus,* and anaerobic gene sequences from the respective analogues which have had their crystal structure solved in the Protein Data Bank. These are summarised in Figure 75A-F.

At a glance there is a high degree of similarity between both the aerobic and anaerobic enzymes, shown by the colour coding in Figure 75, where identical residues are highlighted red and similar residues are highlighted in yellow. The GXGXG motif located just after β1 is present in all these methyltransferases except CbiD, and in the case of $CobL^N$ and CbiE the final glycine residue is not present. An additional motif of GDX, where X is most likely to be a proline, is also observed after β4 in all the methyltransferases, except again in CbiD. Interestingly this shows an aspartic acid backbone interaction in crystal structures which is important for the binding of the methyl donating ligand SAM, Figure 76.

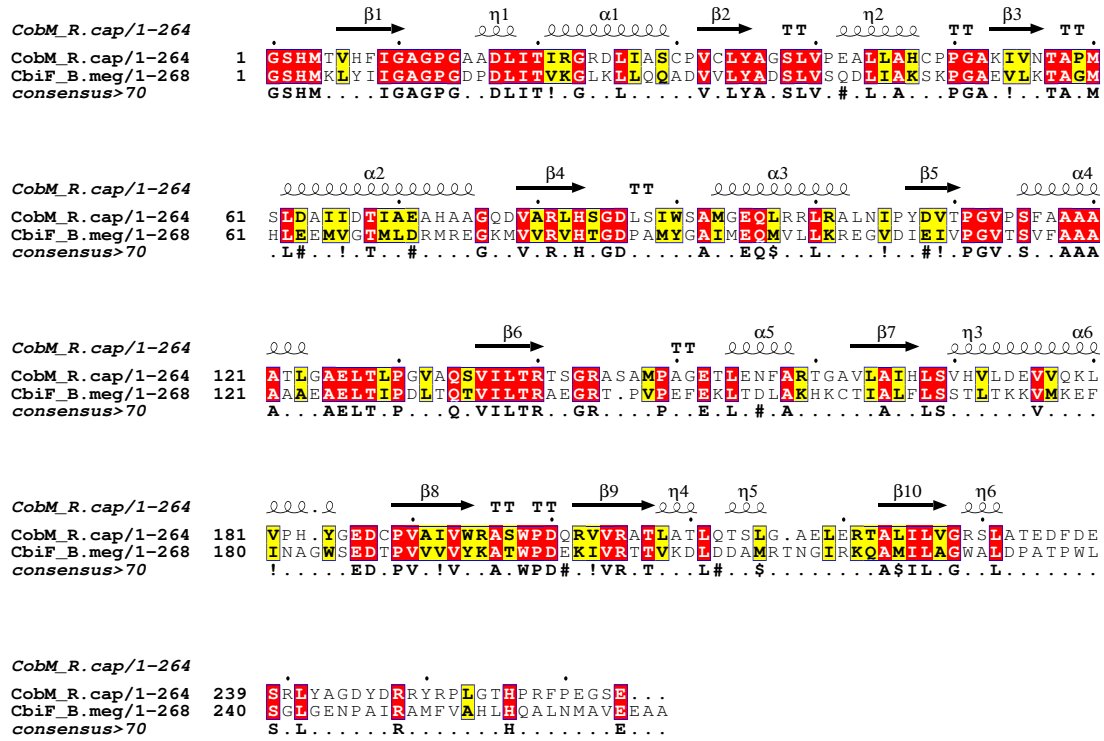
From the sequence alignment of CobF and CbiD, CbiD showed little sequence similarity to CobF. Therefore, CbiD was removed from the final sequence alignment summarised in Figure 75G, as it did not contain either of the conserved motifs present in the other canonical methyltransferases. From this more comprehensive sequence comparison the initial binding motif can be extended from GXGXG where a threonine or serine is located five amino acids after this, located just before α1. A conserved proline is also present after β4, but was not considered relevant for the study as it is likely to be for establishing the link between the two αβα domains present in Class III methyltransferases, as highlighted by **Figure 10** in Chapter 1.


A
SUMT_Sinorhizobium/1-280
SUMT_Sinorhizobium/1-280 1 MIDDLFAGLPALEK.....GSVWLVGAGPGDPGLLTLHAANALRQADVIVHDALVNEDC
CysG_S.typhi/1-253 1 ..ERLFS.EPLDHR.....GEVVLVGAGPGDAGLLTIKGLQQIQQADIVVYDRLVSDDI
CobA_R.cap/1-263 1 MTQILRTSETQAPRRTDHRTGHVTLAGAGPGCADHLTLAVARALAAADVILHDRLVSAEV
consensus>70 ..#.L.....G.V.L.GAGPG....LTL.....AD!!..D.LV..#.
SUMT_Sinorhizobium/1-280 55 LKLARPGAVLEFAGKRGGKPSPKQRDISLRLVELARAGNRVLRLKGGDPFVFGRGGEEAL
CysG_S.typhi/1-253 52 MNLVRRDADRVFVGKRAGYHCVPQEEINQILLREAQKCKRVVRLKGGDPFIFGRGGEELE
CobA_R.cap/1-263 61 LALALAGPQAEVIETGKTGFGAHMPQEEIIALMLGFAQAQGRNVLRLKSGDPGVFGRLGEELD
consensus>70 $.L....A.....GK.....Q.#I...$...A..G..V.RLK.GDP.!FGR.GEE..
SUMT_Sinorhizobium/1-280 115 TLVEHQVPFRIVPGITAGIGGLAYAGIPVTHREVNHAVTFLTGHDSSGLVPDRINWQGIA
CysG_S.typhi/1-253 112 TLCHAGIPFSVVPGITAASGCSAYSGIPLTHRDYAQSVRLVTGHLKTG...GELDWENLA
CobA_R.cap/1-263 121 ALEAAGIGFTVLPGITAASAAAAGLGQSLTERGRNRELRLITGHDVAG..FAEADWAGLA
consensus>70 .L....!.F.!.PGITA.....A..G...T.R.....TGH...G.....#W...A
SUMT_Sinorhizobium/1-280 175 SGSPVIVMYMAMKHIGAITANLIAGGRSPDEPVAFVCNAATPQQAVLETTLARAEADVAA
CysG_S.typhi/1-253 169 AEKQTLVFYMGLNQAATIQEKLIAFGMQADMPVALVENGTSVKQRVVHGVLT..QLGELA
CobA_R.cap/1-263 179 RPGAVAAIYMGKKAARFLQGRLMMHGALATTPVTLVENVSRAEERVIATTLA..DLPAAA
consensus>70 .....YM.....L...G.....PV..V.N.....#.V....L...#....A
SUMT_Sinorhizobium/1-280 235 AGLEPPAIVVGEVVRLRAALDWIGALDGRKLAADPFANRILRNPA
CysG_S.typhi/1-253 227 QQVESPALIIVGRVVALRDKLNWFSNH.....
CobA_R.cap/1-263 237 ARLEGPAVILFGLAPRAAAAQSLEAVI.....
consensus>70 ...E.PA.!..G.....
B
CbiL_M.ther/1-232
CbiL_M.ther/1-232 1 MHGKLIGVGVGPGDSELLTLRAVNVLRSVPVICAPRSSSERESIALSIVEDILTERRDGC
CobI_R.cap/1-245 1 .MGRILCAGLGPGDPDLISVRSDRAIRAATDIAHFRKAG.RAGQARRIVEGMLRPDVVEH
consensus>70 ..G.....G.GPGD.#L...R.....R....I...R....R...A..IVE..L.....
CbiL_M.ther/1-232 61 RILDPV.....FPMTDDRDELESHWDSAARMVAAELEDGRDVAFITLGDPSIYSTFSYLQ
CobI_R.cap/1-245 59 AMEYPVTTELPFDSPDYIEALARFYDEWASRLS.ELSKTRDVVVLCEGDPFLYGSFMHLY
consensus>70 ....PV.....F...D..#.L....D..A.....EL...RDV.....GDP..Y..F..L.
CbiL_M.ther/1-232 116 QRIEDMGFKTEMVPGVTSFTACAATAGRTLVEGDEILLVVP..RVDDRFERVLRDVDACV
CobI_R.cap/1-245 118 TRLQGR.AEVEVIPGIPGMVGCWWATGTPITWGDDVLTVMAGTLPEADLVAHMERSDAIV
consensus>70 .R.#.....E.!PG!.....C....G.....GD#!L.V.....#.....$...DA.V
CbiL_M.ther/1-232 174 IMKTSRHGRRAMEVVESDPRGKDVVSVANCSMDDEVVERGFASGGG...YLATTLVRFRE
CobI_R.cap/1-245 177 VMKTGRNLPKIKRALAVAGRLDAAWLVERGTMPTQRIARLVDVAAEDCPYFAIVLVHGQG
consensus>70 !MKT.R.....R.....V....M..#.!.R.....Y.A..LV....
CbiL_M.ther/1-232
CbiL_M.ther/1-232 231 QS.....
CobI_R.cap/1-245 237 RRPELGDEA
consensus>70 .....

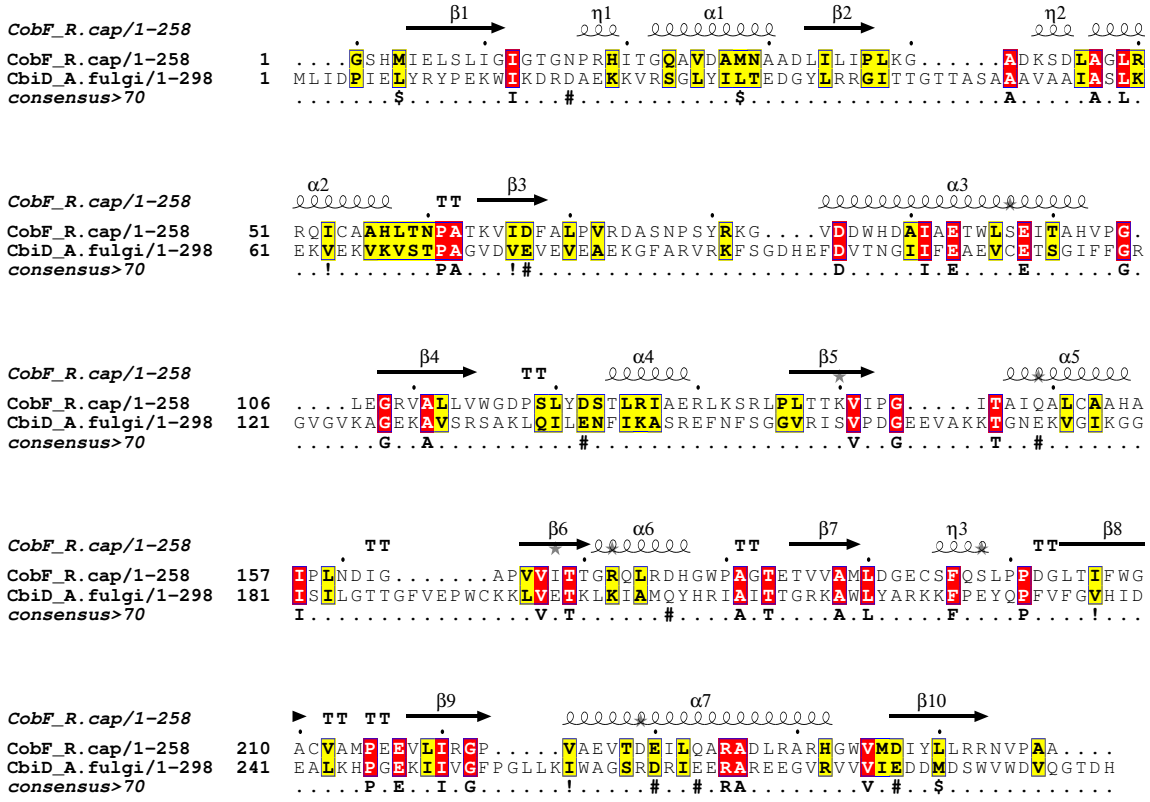
C


CobJ_R.cap/1-245
CobJ_R.cap/1-245 1 MSGWVTVAGLGPGREDLVTPEVTAALAEATDIVGYIPYVARIA.PREGLTLHPTDNRVEL
CbiH_B.meg/1-252 1 MKGKLLVIGFGPGSFEHITQRAREAIQESDMIIGYKTYVELIQGLLTNQQIISTGMTEEV
consensus>70 M.G...V.G.GPG..#.!T.....A..E...I!GY..YV..I..........T....E.
CobJ_R.cap/1-245
CobJ_R.cap/1-245 60 DRATHALEMAAEGRRVVVVSSGDPGVFAMASALFEALEAH..PEHAGTEIRILPGITAML
CbiH_B.meg/1-252 61 SRAQEAVKQAEAGKTVAVISSGDAGVYGMAGLVYEVLIEKGWKKETGVEMEVIPGISAIN
consensus>70 .RA..A...A..G..V.V!SSGD.GV%.MA...%E.L.........G.E..!.PGI.A..
CobJ_R.cap/1-245
CobJ_R.cap/1-245 118 AAAAAGAPLGHDFCAINLSDNLKPFEILEKRLRHAARGDFAMAFYNPRSKSRPHQFTRV
CbiH_B.meg/1-252 121 SCASLLGAPVMHDACTISLSDHLTPWELIEKRIEAAAQADFVVAFYNPKSGRRTRQIVEA
consensus>70 ..A...GAP..HD.C.I.LSD.L.P.E..EKR...AA..DF..AFYNP.S..R..Q....
CobJ_R.cap/1-245
CobJ_R.cap/1-245 178 LEILREECEPGRLILFARAVTTPEQAISVVELRDATPEMADMRTVVLVGNAATRRVGPWV
CbiH_B.meg/1-252 181 QRILLKYRSPDTPVGLVKSAYRDREEVVMTNLKDMLNHEIGMLTTVVVGNSSTFFYDDLM
consensus>70 ..IL.....P...!..........#.!...#L.D.......M.T.V.VGN..T.......
CobJ_R.cap/1-245
CobJ_R.cap/1-245 238 YTPRGVAP....
CbiH_B.meg/1-252 241 ITPRGYQRKYTL
consensus>70 .TPRG.......

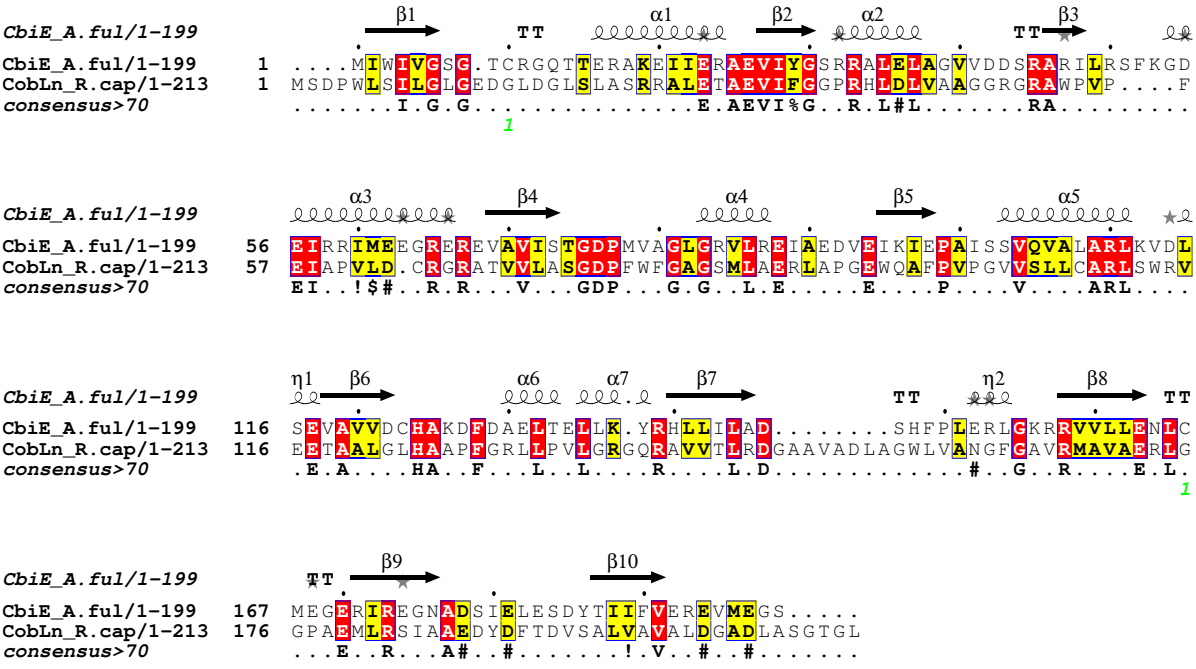
D


CobM_R.cap/1-264
CobM_R.cap/1-264 1 GSHMTVHFIGAGPGAADLITIRGRDLIASCPVCLYAGSLVPEALLAHCPPGAKIVNTAPM
CbiF_B.meg/1-268 1 GSHMKLYIIGAGPGDPDLITVKGLKLLQQADVVLYADSLVSQDLIAKSKPGAEVLKTAGM
consensus>70 GSHM....IGAGPG..DLIT!.G..L.....V.LYA.SLV.#.L.A...PGA.!..TA.M
CobM_R.cap/1-264
CobM_R.cap/1-264 61 SLDAIIDTIAEAHAAGQDVARLHSGDLSIWSAMGEQLRRLRALNIPYDVTPGVPSFAAAA
CbiF_B.meg/1-268 61 HLEEMVGTMLDRMREGKMVVRVHTGDPAMYGAIMEQMVLLKREGVDIEIVPGVTSVFAAA
consensus>70 .L#..!.T..#....G..V.R.H.GD.....A..EQ$..L....!..#!.PGV.S..AAA
CobM_R.cap/1-264
CobM_R.cap/1-264 121 ATLGAELTLPGVAQSVILTRTSGRASAMPAGETLENFARTGAVLAIHLSVHVLDEVVQKL
CbiF_B.meg/1-268 121 AAAEAELTIPDLTQTVILTRAEGRT.PVPEFEKLTDLAKHKCTIALFLSSTLTKKVMKEF
consensus>70 A...AELT.P...Q.VILTR..GR....P..E.L.#.A......A..LS......V....
CobM_R.cap/1-264
CobM_R.cap/1-264 181 VPH.YGEDCPVAIVWRASWPDQRVVRATLATLQTSLG.AELERTALILVGRSLATEDFDE
CbiF_B.meg/1-268 180 INAGWSEDTPVVVVYKATWPDEKIVRTTVKDLDDAMRTNGIRKQAMILAGWALDPATPWL
consensus>70 !.....ED.PV.!V..A.WPD#.!VR.T...L#..$.........A$IL.G..L.......
CobM_R.cap/1-264
CobM_R.cap/1-264 239 SRLYAGDYDRRYRPLGTHPRFPEGSE...
CbiF_B.meg/1-268 240 SGLGENPAIRAMFVAHLHQALNMAVEEAA
consensus>70 S.L......R.......H.......E...

E


CobF_R.cap/1-258
CobF_R.cap/1-258 1 ....GSHMIELSLIGIGTGNPRHITGQAVDAMNAADLILIPLKG.....ADKSDLAGLR
CbiD_A.fulgi/1-298 1 MLIDPIELYRYPEKWIKDRDAEKKVRSGLYILTEDGYLRRGITTGTTASAAAVAAIASLK
consensus>70 .....$.....I...#.....$.....A.....A.L.
CobF_R.cap/1-258 51 RQICAAHLTNPATKVIDFALPVRDASNPSYRKG....VDDWHDAIAETWLSEITAHVPG.
CbiD_A.fulgi/1-298 61 EKVEKVKVSTPAGVDVEVEVEAEKGFARVRKFSGDHEFDVTNGIIFEAEVCETSGIFFGR
consensus>70 ..!.....PA...!#.....D.....I.E....E.....G.
CobF_R.cap/1-258 106 ....LEGRVALLVWGDPSLYDSTLRIAERLKSRLPLTTKVIPG.....ITAIQALCAAHA
CbiD_A.fulgi/1-298 121 GVGVKAGEKAVSRSAKLQILENFIKASREFNFSGGVRISVPDGEEVAKKTGNEKVGIKGG
consensus>70 .....G..A.....#.....V..G.....T..#.....
CobF_R.cap/1-258 157 IPLNDIG.....APVVITTGRQLRDHGWPAGTETVVAMLDGECSFQSLPPDGLTIFWG
CbiD_A.fulgi/1-298 181 ISILGTTGFVEPWCKKLVETKLKIAMQYHRIAITTGRKAWLYARKKFPEYQPFVFGVHID
consensus>70 I.....V.T.....#....A.T....A.L.....F....P....!...
CobF_R.cap/1-258 210 ACVAMPEEVLIRGP.....VAEVTDEILQARADLRARHGWVMDIYLLRRNVPAA....
CbiD_A.fulgi/1-298 241 EALKHPGEKIIVGFPGLLKIWAGSRDRIEERAREEGVRVVVIEDDMDSWVWDVQGTDH
consensus>70 .....P.E..I.G.....!.....#..#.RA.....V.#..$.....

F


CbiE_A.ful/1-199
CbiE_A.ful/1-199 1 ....MIWIVGSG.TCRGQTTERAKEIIERAEVIYGSRRALELAGVVDDSRARILRSFKGD
CobLn_R.cap/1-213 1 MSDPWLSILGLGEDGLDGLSLASRRALETAEVIFGGPRHLDLVAAGGRGRAWPVP....F
consensus>70 .....I.G.G.....E.AEVI%G..R.L#L.....RA.....
CbiE_A.ful/1-199 56 EIRRIMEEGREREVAVISTGDPMVAGLGRVLREIAEDVEIKIEPAISSVQVALARLKVDL
CobLn_R.cap/1-213 57 EIAPVLD.CRGRATVVLASGDPFWFGAGSMLAERLAPGEWQAFPVPGVVSLLCARLSWRV
consensus>70 EI..!$#..R.R...V...GDP...G.G..L.E.....E....P....V....ARL....
CbiE_A.ful/1-199 116 SEVAVVDCHAKDFDAELTELLK.YRHLLILAD.....SHFPLERLGKRRVVLLENLC
CobLn_R.cap/1-213 116 EETAALGLHAAPFGRLLPVLGRGQRAVVTLRDGAAVADLAGWLVANGFGAVRMAVAERLG
consensus>70 .E.A....HA..F...L..L....R....L.D.....#..G..R....E.L.
CbiE_A.ful/1-199 167 MEGERIREGNADSIELESDYTIIFVEREVMEGS.....
CobLn_R.cap/1-213 176 GPAEMLRSIAAEDYDFTDVSALVAVALDGADLASGTGL
consensus>70 ...E..R...A#..#.....!.V..#..#.....

G

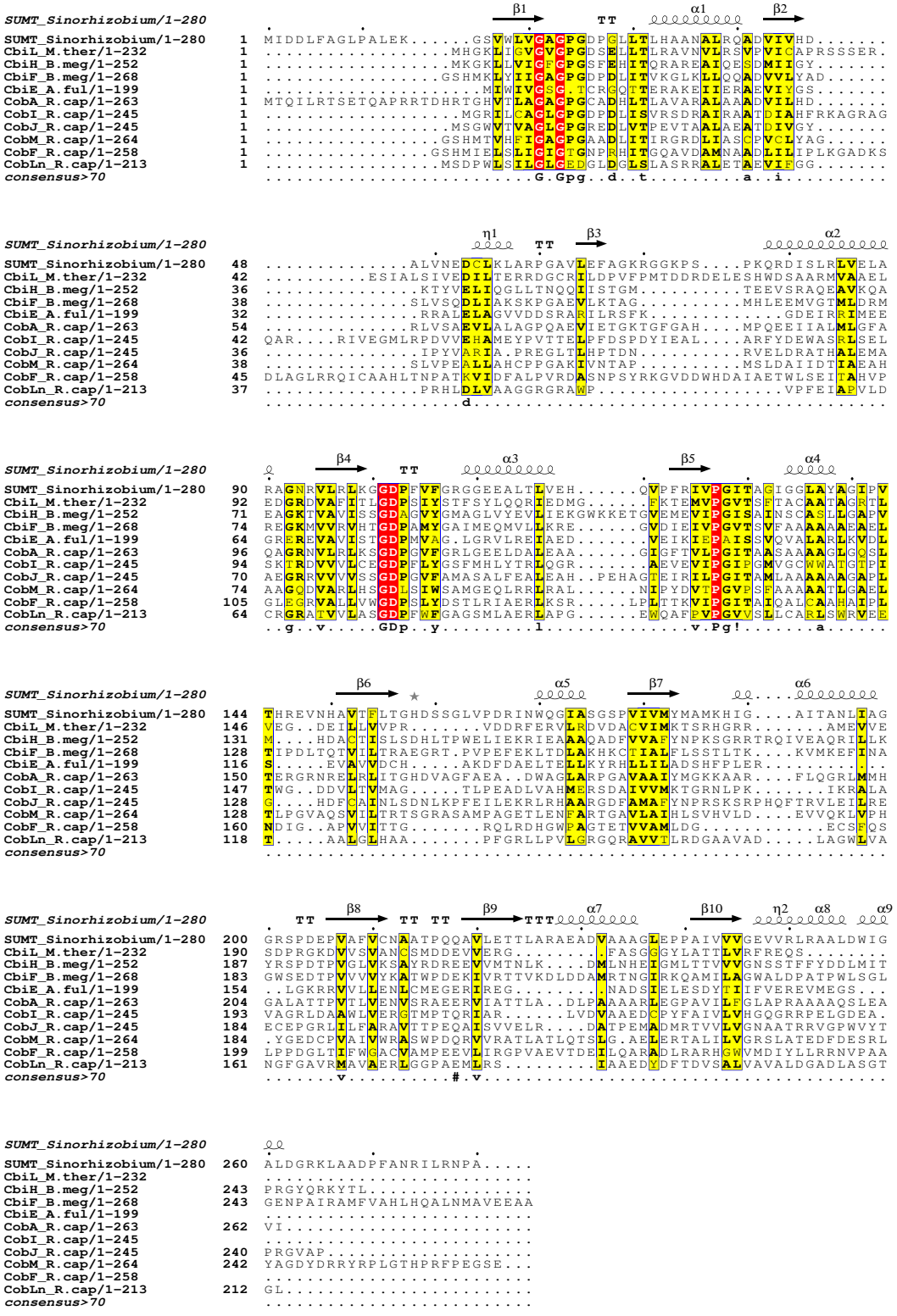


Figure 75: A sequence comparison of the methyltransferase pairs from the aerobic and anaerobic pathways, where red shows identical residues and yellow shows similar residues. A) A sequence comparison of CobA, CysG and SUMT. B) A sequence comparison of CobI and CbiL. C) A sequence comparison of CobJ and CbiH. D) A sequence comparison of CobM and CbiF. E) A sequence comparison of CobF and CbiD. F) A sequence comparison of CobL^{N} and CbiE. G) A comparison of all the aforementioned methyltransferases with CbiD removed. Figure generated using ESPript 3.0 web server[117].

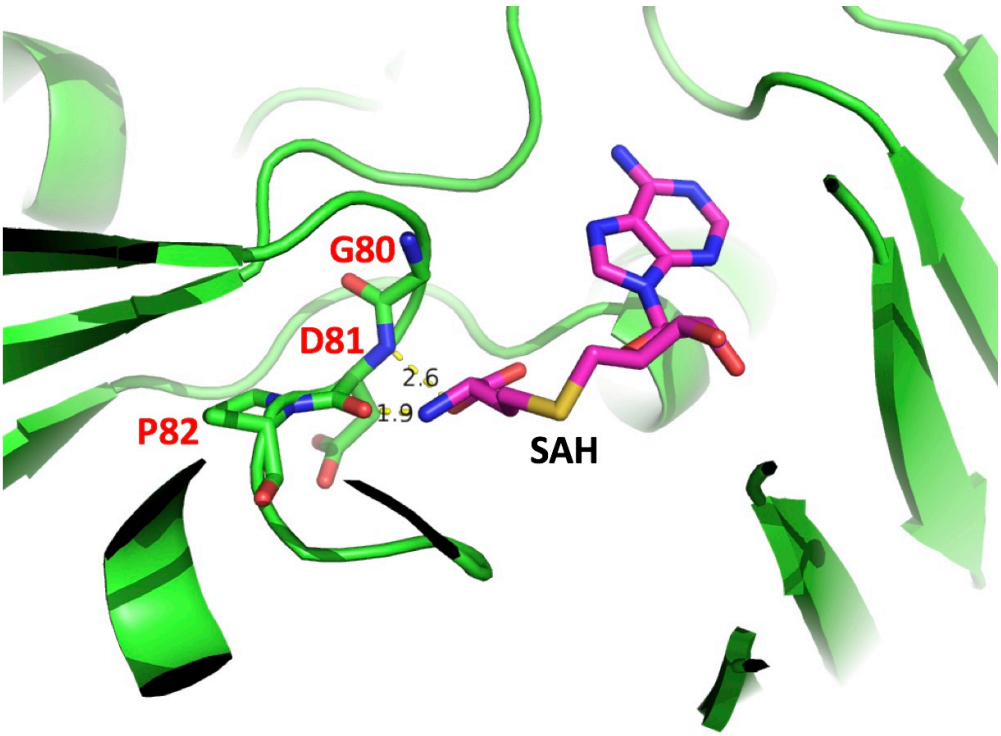


Figure 76: SAH docked in CobJ (3NUT.pdb). The GDP motif is highlighted showing how the conserved aspartic acid interacts with SAH, an unmethylated SAM, through a protein backbone interaction. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

5.3 Docking data of methyltransferases

To test the AutoDock Vina software the docking of HBA to CobH (4FDV.pdb) was trialled given that the pdb file already contains a structure of HBA bound within CobH. The HBA coordinates were stripped from 4FDV.pdb and used as an input for docking. The AutoDock Vina software predicted the experimental structure very closely. It also correctly predicted the HBA position for both binding sites, when the search space was extended and also when the structure of HBA is randomized in Autodock. This trial run gave a binding affinity of -13.7kcal/mol.

An unpublished structure of CbiET, solved by the Pickersgill group, containing Cobalt-precorrin-6 ligand present was also trialled. Cobalt-precorrin-6B binds to the C-terminal domain in essentially the same way as PC6B in the non-canonical methyltransferase $CobL^{c}$ in which *in silico* docking studies gave a binding affinity of -8.5kcal/mol. This is believed to provide a more realistic bench mark for less tightly binding precorrin ligands.

5.3.1 Docking of PC4 (product) to CobJ (3NUT.pdb)

Initial docking of PC4 showed that the cleft was too tight for PC4 to penetrate into the right depth, so NMA was undertaken using the ElNemo web server. Eleven structures in this mode were generated with varying degrees of cleft opening. The widest cleft gave generated models with the best binding affinity energies between -7.5kcal/mol and -7.2kcal/mol. These results showed that the degree of cleft opening does play a crucial role in binding. Of the nine docked structures generated a summary of the distance between the C17 methylation site on PC4 and sulphur atom of SAH, which binds the methyl group in SAM, are shown in Table 13.

Docking model of PC4 to CobJ	Distance between C17 and S atoms (Å)
1	4.9
2	4.8
3	8.2
4	8.6
5	6.6
6	10.8
7	11.1
8	9.5
9	9.9

Table 13: Summary table of docking models obtained from Autodock Vina for the protein CobJ with ligand PC4 and the distance measured between methylation site and sulphur atom of SAH.

Models which gave a distance greater than 7Å between SAH and ligand were eliminated as this distance is roughly two hydrogen bond lengths and this would normally be occupied by a methyl group, which is available for transfer within the active site. This therefore left models 1, 2 and 5 upon which the polar interaction of the acetic and propanoic decorations of the PC4 corrin ring were identified. The amino acids involved in these interactions, hydrogen bond distances and whether they came about via an R-group or protein backbone interaction are summarised in Table 14. In a few cases this interaction came about via an amino acid on a second protein chain so was underlined to signify this. From this summary models 1 & 2 were almost identical and therefore the most probable solution generated. The key amino acid interactions from model 1 are shown in Figure 77.

Location of PC4 decoration	Model 1 Amino acid interactions and distance (Å)	Model 2 Amino acid interactions and distance (Å)	Model 5 Amino acid interactions and distance (Å)
C1 **(ethanone)**	**Backbone** S137- 3.4 **R-group** S137- 3.0 N164-3.4	**Backbone** S137- 3.4 **R-group** S137- 3.0 N164-3.4	-
C2 **(acetic acid)**	**Backbone** S137- 3.4 N139-3.3 **R-group** N139-3.2	**Backbone** S137-3.0 N139-3.3 **R-group** N139-3.2	-
C3 **(propanoic acid)**	**R-group** S169-3.0	**Backbone** K168-3.3 **R-group** S169-3.0	-
C7 **(acetic acid)**	-	-	-
C8 **(propanoic acid)**	-	-	**Backbone** S137-3.4 **R-group** N139-3.2 S167-2.7 S169-3.0
C12 **(acetic acid)**	**Backbone** N55-3.1 **R-group** N55-3.2 S80-3.0	**R-group** S80-2.8	**R-group** D138-3.5 R241-3.2
C13 **(propanoic acid)**	**Backbone** S80-3.2 **R-group** S79-3.2 S80-3.0	**Backbone** S79-3.2 S80-3.3 **R-group** S80-2.9	**Backbone** A87-3.0 **R-group** R241-2.9
C17 **(propanoic acid)**	**R-group** R241-3.1	**R-group** R241-3.0	**Backbone** S80-2.9 **R-group** S79-3.0 S80-2.8
C18 **(acetic acid)**	**-**	-	**R-group** N55-3.1 S80-2.8

Table 14: From the most feasible model predictions for PC4 to CobJ, the carbon, upon which the carboxylate decoration is located to the precorrin ring is summarised, showing the amino acid interactions involved.

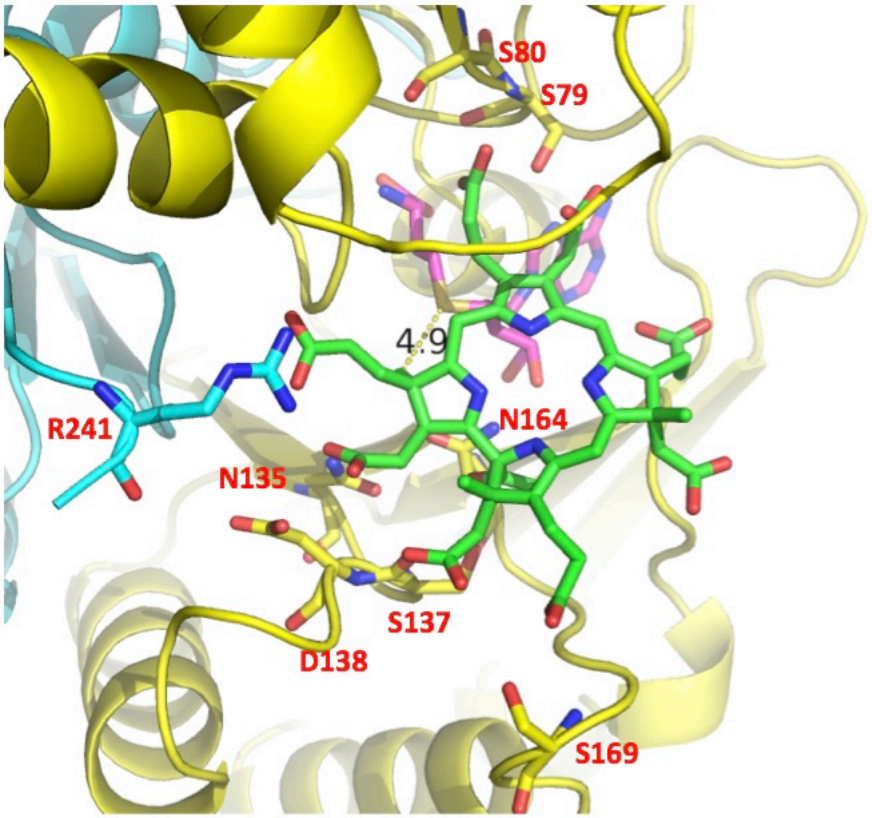


Figure 77: The amino acids which are involved in docking interactions of PC4 to CobJ (3NUT.pdb). SAH is coloured purple, PC4 is coloured green. Chain A of CobJ is coloured yellow and chain B is coloured blue, which shows R241 contributes from the other chain. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

5.3.2 Docking of PC4 (substrate) and PC5 (product) to CobM (3NDC.pdb)

Docking of both the substrate PC4 and product PC5 were trialled with CobM. The following residues in CobM were permitted flexibility: L36, S81, R137, R141, L165, R218 and Q131 (where Q131 was located on chain A, while the rest are located on chain B). The following binding affinity energies and distance between C11 and SAH are summarised in Table 15 and Table 16.

Docking model of PC4 to CobM	Binding affinity (kcal/mol)	Distance from best model rmsd (I.B)	Distance from best model rmsd (U.B)	Distance between C11 and S atoms (Å)
1	-7.6	0.000	0.000	5.9
2	-7.6	2.054	6.067	7.1
3	-7.4	1.515	2.014	6.2
4	-7.3	1.742	2.398	6.6
5	-7.3	2.053	6.137	8.2
6	-7.1	2.087	6.346	6.3
7	-7.1	1.395	1.935	5.9
8	-7.0	3.183	8.682	8.1
9	-7.0	1.911	6.432	7.1

Table 15: Summary table of docking models obtained from Autodock Vina for the protein CobM with substrate PC4. Binding affinity, distance from best model and the distance measured between methylation site and sulphur atom of SAH are all shown.

Docking model of PC5 to CobM	Binding affinity (kcal/mol)	Distance from best model rmsd (I.B)	Distance from best model rmsd (U.B)	Distance between C11 and S atoms (Å)
1	-8.1	0.000	0.000	8.2
2	-8.0	1.884	7.146	8.3
3	-8.0	1.319	1.818	8.3
4	-7.9	2.136	7.346	7.3
5	-7.7	1.658	2.167	8.4
6	-7.5	1.875	6.354	4.9
7	-7.4	2.261	6.245	7.7
8	-7.3	2.189	6.153	5.3
9	-7.3	1.774	6.488	7.3

Table 16: Summary table of docking models obtained from Autodock Vina for the protein CobM with product PC5. Binding affinity, distance from best model and the distance measured between methylation site and sulphur atom of SAH are all shown.

The NMA technique was applied to CobM to check if different degrees of cleft opening would facilitate docking. Both a more open and tighter cleft opening using NMA did not help to correctly dock PC5 to CobM. All 9 different cleft openings contain one or two correct ligand conformations, but C11 is still quite far away from the S-atom of SAH. In a similar fashion to CobJ above, a summary of the binding interactions to corrin ring decoration of the ligands PC4 and PC5 is summarised in Table 17 and Table 18 respectively. From these tables, model 6 from PC4 docking and model 1 from PC5 docking could be paired together based on the similarities of the respective binding energies, distance measured and binding interactions. The key amino acid interactions from PC4 model 1 and PC5 model 6 are shown in Figure 78.

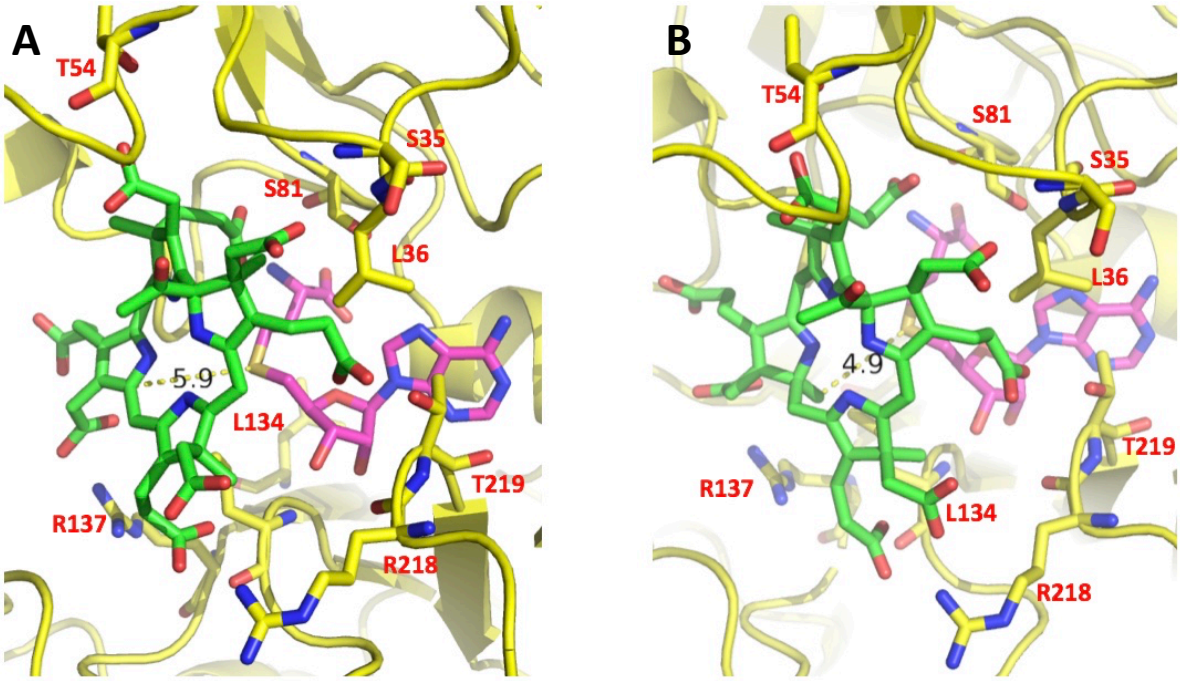


Figure 78: A) The amino acids which are involved in docking interactions of PC4 to CobM (3NDC.pdb). B) The amino acids which are involved in docking interactions of PC5 to CobM. SAH is coloured purple, PC4 is coloured green and CobM is coloured yellow. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

Location of PC4 decoration	Model 1 Amino acid interactions and distance (Å)	Model 2 Amino acid interactions and distance (Å)	Model 3 Amino acid interactions and distance (Å)	Model 4 Amino acid interactions and distance (Å)	Model 6 Amino acid interactions and distance (Å)	Model 7 Amino acid interactions and distance (Å)	Model 9 Amino acid interactions and distance (Å)
C1	-	**Backbone** S88-3.4	-	-	**Other** SAH-3.0	-	-
C2	**Backbone** S35-3.0 L36-3.3	**Backbone** A55-3.5 **R-group** T54-2.8	**Backbone** A33-2.8	**Backbone** S35-3.0 **R-group** S35-3.2	-	-	-
C3	**R-group** T219-2.9	**Backbone** M57-3.3 L59-3.2 **R-group** E92-3.3	-	-	-	-	-
C7	-	**R-group** E92-3.4	-	-	**R-group** S81-2.9		**R-group** S81-2.6
C8	**Backbone** L165-3.3 **R-group** R218-2.9	**Backbone** A142-3.3 **R-group** R137-3.0	**R-group** H168-3.2 R218-3.1	**Backbone** G140-3.5 **R-group** H168-3.1 R218-3.3	**R-group** H80-3.2 S81-3.3 **Other** SAH-3.4	**R-group** H168-3.4 R218-3.1	**R-group** T54-3.1 R78-3.1
C12	**R-group** R137-3.1	**R-group** R218-2.9	-	**Backbone** A142-3.2 **R-group** R137-2.7	-	**R-group** R137-3.2	**Backbone** W87-3.3
C13	-	**Backbone** R218-3.2 **R-group** R218-3.2	**R-group** Q131-2.5 R137-1.6	**R-group** R137-2.6	**R-group** Q131-3.0 R137-2.5	**R-group** E123-3.5	**Backbone** A142-3.2 **R-group** R137-2.2
C17	**Backbone** S81-3.1	**Backbone** S35-3.2 L36-3.2 **R-group** S35-3.2	**Backbone** S-81-3.1	**Backbone** S88-3.3 **R-group** H80-2.9	**R-group** R218-2.0	**Backbone** S81-3.2 **R-group** R218-2.9	**R-group** H168-3.2
C18	**R-group** T54-2.9	**Backbone** A33-2.9 **R-group** S81-2.6	-	**Backbone** A33-3.4	**R-group** H168-3.3 R218-3.0 **Other** SAH-2.9	-	**Other** SAH-2.8

Table 17: From the most feasible model predictions for PC4 to CobM, the carbon, upon which the carboxylate decoration is located to the precorrin ring is summarised, showing the amino acid interactions involved. Interactions occurring from the other protein chain are underlined.

Location of PC5 decoration	Model 6 Amino acid interactions and distance (Å)	Model 8 Amino acid interactions and distance (Å)
C1 **(ethanone)**	-	-
C2 **(acetic acid)**	**Backbone** S35-3.3 L36-3.1	**R-group** E92-3.4
C3 **(propanoic acid)**	**R-group** T219-2.7	**Backbone** T54-3.4 A55-3.4 **R-group** T54-2.8 R78- 3.1
C7 **(acetic acid)**	**R-group** R218-2.7	-
C8 **(propanoic acid)**	**Backbone** L135-3.3 **R-group** R218-2.9	**R-group** R218-2.9 **Other** SAH-3.0
C12 **(acetic acid)**	**R-group** R137-3.0	**R-group** S80-2.8
C13 **(propanoic acid)**	**-**	**-**
C17 **(propanoic acid)**	**Backbone** S81-3.1	-
C18 **(acetic acid)**	**R-group** T54-2.8	**-**

Table 18: From the most feasible model predictions for PC5 to CobM, the carbon, upon which the carboxylate decoration is located to the precorrin ring is summarised, showing the amino acid interactions involved.

5.3.3 Docking of PC6A (substrate) to CobF (3ND1.pdb)

As above, summaries of CobF docking are outlined below in Table 19 and Table 20. The following residues in CobF were made flexible: I159 (A-chain), D39 (B-chain), Y118 (B-chain), S120 (B-chain), R169 (B-chain), R236 (B-chain), M243 (B-chain). When SAH was set to flexible its position became less stable when compared to the other methyltransferases. In general, the docking of PC5 to CobF gave fairly unsatisfactory results with only one potentially feasible model, showing much less consistency. The key amino acid interactions from model 2 are shown in Figure 79.

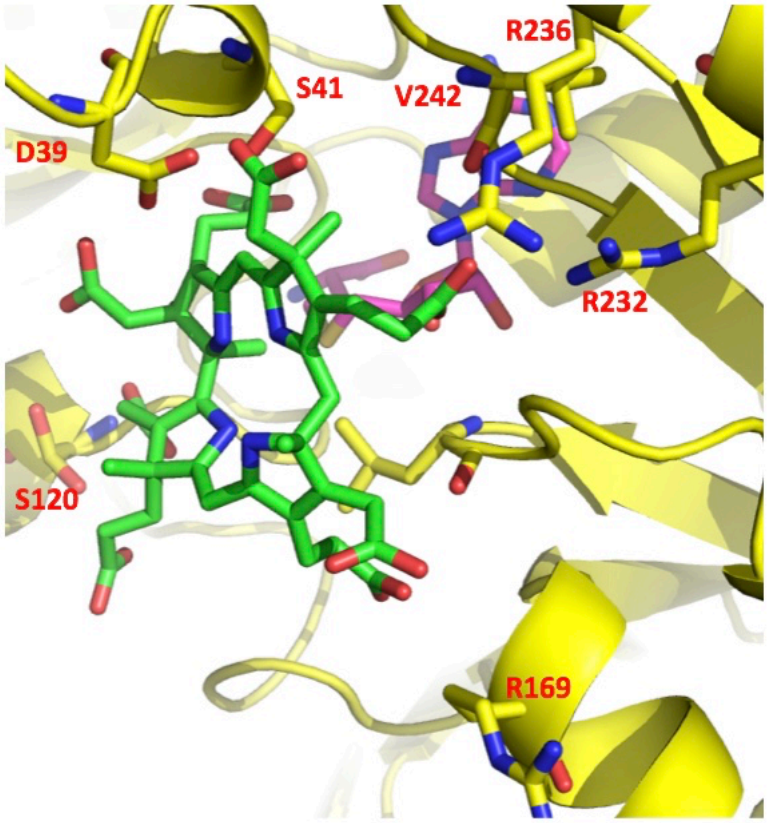


Figure 79: The amino acids which are involved in docking interactions of PC6A to CobF (3ND1.pdb). SAH is coloured purple, PC4 is coloured green and CobF is coloured yellow. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

Docking model of PC5 to CobF	Binding affinity (kcal/mol)	Distance from best model rmsd (I.B)	Distance from best model rmsd (U.B)	Distance between C11 and S atoms (Å)
1	-8.3	0.000	0.000	7.0
2	-8.2	2.024	6.450	5.3
3	-7.7	3.336	8.757	10.0
4	-7.6	3.021	6.622	7.0
5	-7.5	1.527	2.045	10.0
6	-7.4	2.215	6.415	10.1
7	-7.4	3.039	7.312	10.1
8	-7.4	1.677	5.976	8.0
9	-7.4	2.051	6.659	11.4

Table 19: Summary table of docking models obtained from Autodock Vina for the protein CobF with substrate PC5. Binding affinity, distance from best model and the distance measured between methylation site and sulphur atom of SAH are all shown.

Location of PC5 decoration	**Model 2** Amino acid interactions and distance (Å)
C2 (acetic acid)	-
C3 (propanoic acid)	-
C7 (acetic acid)	**R-group** D39-1.8 S41-3.0
C8 (propanoic acid)	**Backbone** V242-3.3 **R-group** R232-3.0 R236-3.5
C12 (acetic acid)	**R-group** R169-3.1
C13 (propanoic acid)	**Backbone** L187-3.4 **R-group** R169-3.3
C17 (propanoic acid)	**R-group** S120-2.8
C18 (acetic acid)	**Backbone** S120-3.0 **R-group** S120-2.9

Table 20: From the most feasible model predictions for PC5 to CobF, the carbon, upon which the carboxylate decoration is located to the precorrin ring is summarised, showing the amino acid interactions involved.

5.3.4 Docking of PC6 (substrate) to CbiE (2BB3.pdb)

CbiE is the anaerobic equivalent of the N-terminal domain of CobL. The docking of PC6B to CbiE is summarised below in Table 21 and Table 22. The docking models one and two are almost identical when comparing binding affinities, distance between C5 and SAH and ligand interactions. H129 can also be seen almost directly below the corrin ring and would likely be involved in co-ordinating cobalt in the anaerobic equivalent Co-PC6B. The key amino acid interactions from model 1 are shown in Figure 80.

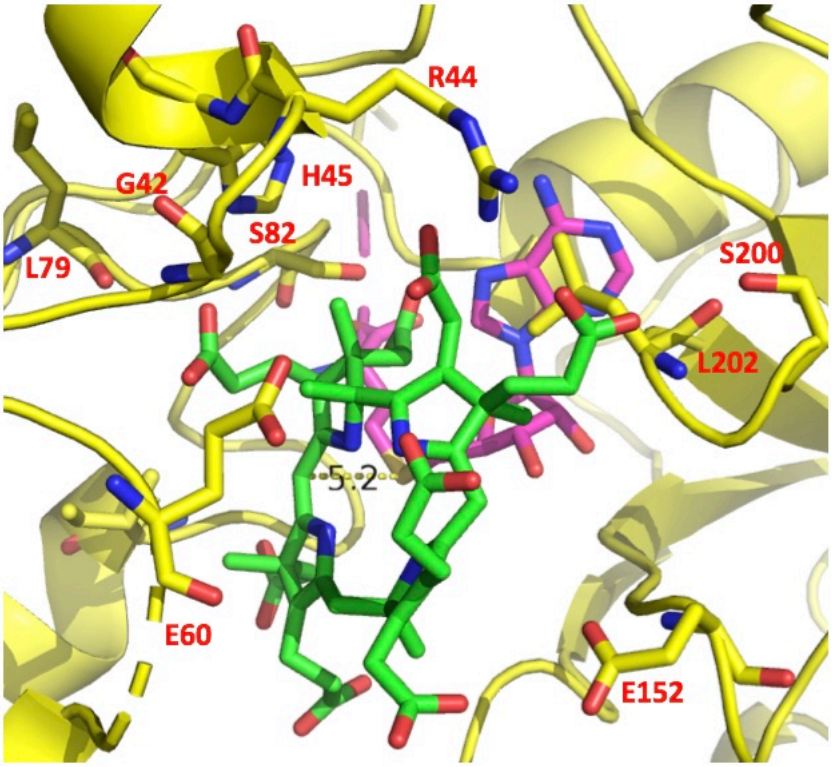


Figure 80: The amino acids which are involved in docking interactions of PC6B to CbiE (unpublished). SAH is coloured purple, PC4 is coloured green and CbiE is coloured yellow. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

Docking model of PC6B to CbiE	Binding affinity (kcal/mol)	Distance from best model rmsd (I.B)	Distance from best model rmsd (U.B)	Distance between C5 and S atoms (Å)
1	-7.5	0.000	0.000	5.2
2	-7.5	1.646	2.545	5.2
3	-7.4	2.443	7.626	9.2
4	-7.3	1.667	2.129	5.0
5	-7.1	2.774	4.161	7.2
6	-7.1	2.207	9.583	7.8
7	-7.1	1.950	9.545	7.8
8	-7.0	2.586	3.720	9.8
9	-7.0	2.113	8.099	10.2

Table 21: Summary table of docking models obtained from Autodock Vina for the protein CbiE with ligand PC6B. Binding affinity, distance from best model and the distance measured between methylation site and sulphur atom of SAH are all shown.

Location of PC6B decoration	Model 1 Amino acid interactions and distance (Å)	Model 2 Amino acid interactions and distance (Å)	Model 4 Amino acid interactions and distance (Å)
C2 (acetic acid)	**R-group** S82-3.2 **Other** SAH-3.2	**R-group** S82-3.2 **Other** SAH-3.2	**R-group** S82-3.2 **Other** SAH-3.3
C3 (propanoic acid)	**Backbone** G42-3.1 L80-3.4 **R-group** H45-3.0	**Backbone** G42-3.1 L80-3.4 **R-group** H45-3.0	**Backbone** S82-3.3 Other SAH-3.1
C7 (acetic acid)	**Backbone** I90-3.3	**Backbone** I90-3.4	-
C8 (propanoic acid)	**-**	**Backbone** H129-3.3 **R-group** H129-3.0	-
C12 (acetic acid)	**R-group** E152-3.4	**-**	-
C13 (propanoic acid)	**R-group** E60-3.4	**R-group** E60-3.5	-
C17 (propanoic acid)	**Backbone** L202-3.2 **R-group** R44-2.9 S199-3.4	**R-group** R44-3.2	**R-group** R44-3.0
C18 (acetic acid)	**R-group** R44-2.9	**R-group** R44-3.0	**R-group** R44-3.2

Table 22: From the most feasible model predictions for PC6B to CbiE, the carbon, upon which the carboxylate decoration is located to the precorrin ring is summarised, showing the amino acid interactions involved.

5.3.5 Summary of docking interactions

It seemed appropriate to summarise the location of the interacting amino acid residues in the structure of the protein given that the orientation of the precorrin ligand changes upon binding depending on the location of the methylation site and docking orientation of the corrin ring. Due to variation observed in the number of alpha helices crystal structures and sequence alignments in Figure 75, a summary of the amino acid locations was based upon the beta-strands present in Table 23, as each methyltransferase contains ten beta strands. CobJ does not quite fit this trend, however it has already been suggested that an additional ancillary reaction may account for this deviation where it also carries out the ring contraction and removal of C20. This summary begins to narrow down the loop regions which are involved in precorrin ligand docking and variations that dictate specificity. All the methyltransferases appear to contain a serine residue post β4, which is also located next to the conserved GDX motif described in section 5.2. In all four enzymes at least one arginine is present and important for ligand docking.

Cob enzymes	**Location of interaction on secondary structure of enzyme**					
	Post β2	**Post β3**	**Post β4**	**Post β6**	**Post β7**	**Post β10**
CobJ	-	N55	S79 S80	S137	N164 S169	R241
CobM	S35 L36	T54	S81	L135 R137	L165	R218 T219
CobF	D39 S41	-	S120 H129	R169	L187	R232 R236 V242
CbiE	G42 R44 H45	E60	L80 S82 I90	H129	E152	S200 L202

Table 23: A summary of where the interacting amino acids are located in each methyltransferase

To further explore the decoration pattern of precorrin binding docking interactions of the pyrrole rings A-D are summarised in Table 24. This table shows that all sides of the precorrin ring are involved in docking, but to varying extents. CobM has two edges interacting with one residue, while CobJ and CobF have only one edge in which this occurs. CbiE shows a large number of binding

interactions across two edges via at least four amino acids. Overall each methyltransferase has at least one edge interacting with at least four amino acid residues.

Cob enzyme	Corrin ring edge based upon which pyrrole rings interactions occur along			
	AB (C3/C7)	**BC (C8/C12)**	**CD (C13/C17)**	**DA (C18/C1/C2)**
CobJ	S169	N55 S80	S79 S80	S137 N139 N164 R241
CobM	T219	L135 R137 L165 R218	S81	S35 L36 T54
CobF	D39 S41	R169 R232 R236 V242	S120 R169 L169	S120
CbiE	G42 H45 L80 I90	H129 E152	R44 E60 S199 L202	R44 S82

Table 24: A summary of which amino acids interact with each pyrrole ring for different precorrin ligands in CobJ, CobM, CobF and CbiE.

Using the most probable docking models for each methyltransferase and keeping the orientation of SAH the same allowed the precorrin pyrrole sequence and decoration to be summarised in Table 25. In the case of CobJ and CbiE the pyrrole docking was anti-clockwise (D→A) whereas the pyrrole docking in CobM and CobF was clockwise (A→D). This in turn changes the decoration sequence of the precorrin ligands, which may drive specificity.

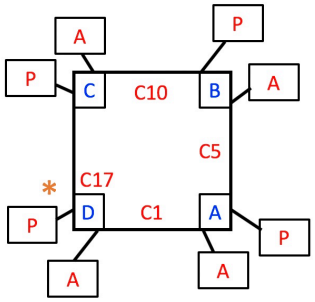
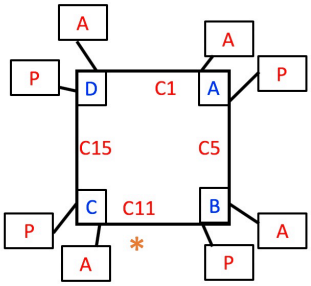
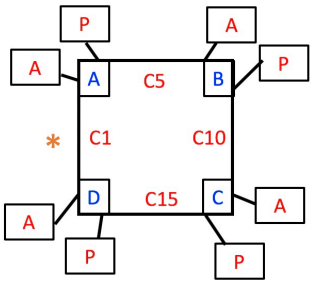
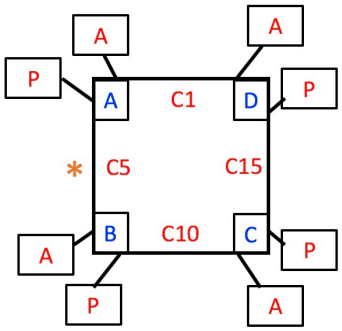
Enzyme	Docking orientation	Edges of precorrin ligand read in a clockwise fashion	Clockwise decoration sequence
CobJ (methylates C17)		DC-CB-BA-AD	PP-AP-AP-AA
CobM (methylates C11)		BC-CD-DA-AB	PA-PP-AA-PA
CobF (methylates C1)		DA-AB-BC-CD	AA-PA-PA-PP
CbiE (methylates C5)		BA-AD-DC-CB	AP-AA-PA-PP

Table 25: A summary of the docking orientation and decoration sequence of precorrin ligands when the architecture of the methyltransferase and SAM/SAH binding site is kept fixed.

5.4 Exploring the β-loop sequences and precorrin ligand decoration patterns

Due to the lack of conserved residues in these loop regions, precorrin specificity may be dictated more by steric factors based on the decoration sequence of the precorrin ring. For example, an acetic acid group is shorter and would therefore allow a longer amino acid R-group to be present, which would otherwise clash with a longer propanoic chain. Therefore, these regions could be assigned to specific pyrrole interactions. Given the methyltransferase tertiary structure and SAM/SAH binding orientation is similar and can remain fixed, a more appropriate method to assign these loop regions could be to use compass descriptions, due to precorrin docking orientation changing. This is summarised in Table 26 where the only consistency is observed in the post β7 region which appears to always interact with the south-east pyrrole group upon ligand docking. The post β4 loop appears to interact with the west side of the precorrin ligand upon docking.

Cob enzymes	**Pyrrole ring upon which docking occurs and description of docking orientation using compass points**					
	Post β2	**Post β3**	**Post β4**	**Post β6**	**Post β7**	**Post β10**
CobJ	-	C-ring NW	C-ring NW	A-ring SE	A-ring SE	D-ring SW
CobM	A-ring NE	D-ring NW	D-ring NW	B-ring & C-ring S	B-ring SE	C-ring SE
CobF	B-ring NE	-	D-ring SW	C-ring SE	C-ring SE	B-ring NE
CbiE	A-ring & D-ring N	C-ring SE	A-ring & B-ring W	B-ring SW	C-ring SE	D-ring NE

Table 26: Table summarising which pyrrole group of the precorrin these loop regions interact with and a compass description of how the ligand is docked (e.g. NW=North-West).

Based on the binding residues described earlier in Table 24 the amino acid sequence ±5 residues upstream and downstream were isolated. These sequences were grouped based on where they are located in the tertiary structure and aligned using Jalview[118]. These were aligned to see if any similarities or conserved residues were present which is summarised in Figure 81. The two most

similar regions were post β4 and post β6 loop regions. The post β4 region contained the aforementioned GDX motif, but the interacting serine residue occurs post this in CobF rather than before it in the others. The post β6 loop contained at least one basic residue and either a serine or threonine. None of the other loops appeared to contain any significant features.

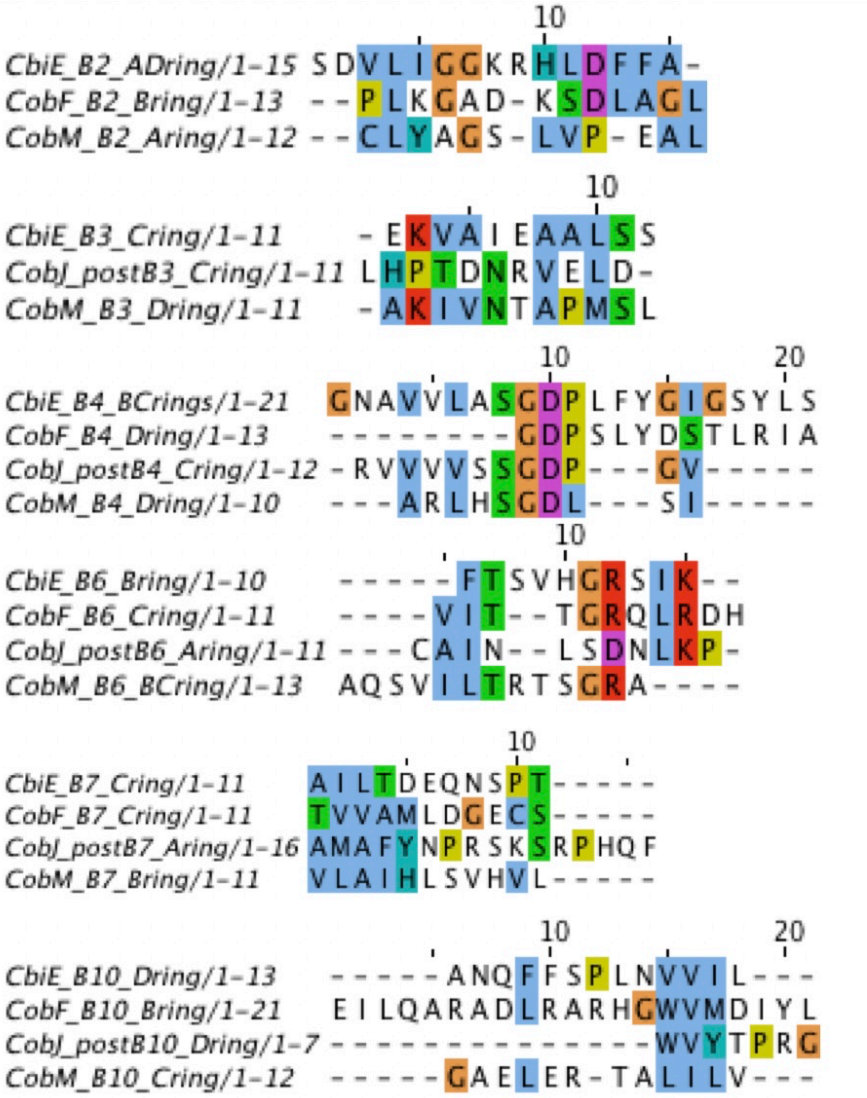


Figure 81: A sequence alignment of post β-strand loop regions across the methyltransferases. The sequences were isolated by taking the residues ±5 residues up-and down-stream of interacting amino acids identified during docking studies. Blue signifies hydrophobic residues, orange signifies a glycine, yellow signifies a proline, green signifies a threonine or serine, red signifies an arginine or lysine, purple signifies an aspartate and teal signifies an aromatic residue.

5.5 Discussion and conclusions

5.5.1 Sequence analyses

The sequence analyses of the canonical methyltransferases did not show any obvious conserved amino acid residues which may be involved in precorrin binding or residues which are important in differentiating between the different precorrin intermediates. Despite using only two sequences for a comparison this was undertaken between enzymes from distinctly different respiratory pathways which carry other identical chemical modifications. Structurally these enzymes are very similar and so it comes as no surprise at the good sequence similarity (yellow) and identity (red) observed in Figure 75A-E. The only pair that did not match well were CobF and CbiD, however this is not surprising given Warren et al, 2002 describes them as genetically distinct due to the nature of the ancillary reaction they carry out, where CobF removes acetic acid at C1 and CbiD removes acetaldehyde[2].

For this reason, CbiD was removed from the more comprehensive sequence analysis shown in Figure 75G between all the canonical methyltransferases of the aerobic and anaerobic pathway. Two sequence motifs are observed, the first of which is GXGXG located after β1. This is supported by Schubert et al, 2003 as a means of identifying this class of methyltransferase, despite not actually being involved in SAM binding[54]. The final glycine residue in this sequence is not fully conserved but given it is not essential for SAM binding, some flexibility may be permitted. Also, six residues upstream of this motif, there appears to be a threonine or serine which is located at the start α1. The second conserved motif present is GDX located after β4, where X is typically a proline. The aspartic acid residue is clearly conserved as structurally it can be observed interacting directly with SAH in crystal structures and therefore essential for SAM binding. Interestingly this is brought about via backbone interactions of the two adjoining amide linkages, rather than the R-group of aspartic acid. This might lead one to speculate that the acetic acid R-group is important for a further interaction. However, upon exploring the structures further there does not structurally appear to be an apparent dependence for this as it is located between the dimer interface and typically interacts

with a water molecule bridged to α5-helix on the second protein chain and therefore is possibly involved in dimer formation but is clearly not solely responsible for this. However this has not been explored using mutagenesis.

5.5.2 Docking analysis

Initial docking trials using CobH were encouraging as the software was able to correctly identify the ligand binding site and dock HBA correctly on to a known enzyme-ligand complex. The binding affinity energies for CobH was quite high at -13.7kcal/mol. However, CobH is known to have a very high affinity for HBA. A second trial of $CobL^c$ *in silico* docking studies, in which predicted docking was matched to the known ligand docking of CbiET, gave a more reasonable binding affinity of -8.5 kcal/mol providing a more realistic benchmark for less tightly bound precorrin ligands in the methyltransferases. The subsequent *in silico* docking studies of precorrin ligands into different methyltransferases all gave similar binding affinities around -7.5kcal/mol in which the SAH ligand was docked correctly and in an identical orientation across all the methyltransferases.

CobJ docking was remarkable in that the binding site of PC4 had very few binding interactions from arginines and lysines, despite them being dotted around the precorrin cleft. These are in fact not well conserved apart from R241 which interacts with the PC4 pyrrole ring A. This would explain why such an extreme C-terminal residue is seen in the electron density map of the crystal structure. In the homodimer this R241 interaction occurs from the other protein chain and is important in substrate binding and orientation. Unpublished work from the Pickersgill group generated *in vitro* CobJ mutants where it was shown that the methylation and ring contraction reactions are in fact uncoupled and their activity is dependent on interactions contributed from each protein chain.

CobM showed a relatively open substrate-binding cleft. Important contacts defining the binding mode are conserved R141 and semi-conserved R/L218 which bind to the C- and D-pyrrole ring groups via basic residues. Conservation is greatest around the SAM-binding site and C-ring, which is in agreement with the need to position C11 for effective catalysis.

CobF gave only one feasible docking model of PC5 based on the distance involved for a methylation reaction to occur. A number of arginine residues are involved in binding interactions including R169 which binds to the C-ring and R232 and R236 which bind to the B-ring. The lack of formally positively charged residues in the proximity of the A-ring could be related to the requirements of catalysis where protonation of an aspartate by water could lead to generation of a hydroxy-group to attack the C1 acetyl substituent with subsequent methylation from the other side of the ring. D39 together with the A-ring acetate may activate a water molecule to attack the C1 carbonyl-carbon, while conserved D119 on the other side of the ring, close to SAH, may also be catalytic.

CbiE docking can in fact be confirmed by cross-checking with the unpublished structure of CbiET solved by the Pickersgill group. This structure confirms that H129 is involved in co-ordinating to cobalt in Co-PC6B, which would not be needed in $CobL^N$ of the aerobic pathway, but it can be assumed the remaining interactions would be comparable. This structure and docking confirm that the majority of interactions occur to the A-ring to guide orientation for C5 methylation through the basic residue R44, whereas the C- and D-ring are left more exposed, supporting the binding affinities recorded for less tightly bound ligands.

Each docking attempt generated multiple models and by carefully inspecting and comparing these, including the distance between the sulphur on SAH with the methylation site, the most appropriate model could be selected. This allowed amino acid interactions to be summarised based on their position in the protein structure. In the case of CobJ and CbiE the precorrin ligand docked in an anticlockwise fashion with respect to the pyrrole groups, while in CobM and CobJ docking occurred in a clockwise fashion, as summarised in Figure 82. Therefore, having suggested in chapter 4 that CobM may be the culprit for additional C5 methylation, observed in the CobK-PC6B crystal structure, CobJ may in fact be a more appropriate suspect. This is due to the fact that both precorrin ligands bind in an anticlockwise fashion rather than a clockwise fashion seen in CobM docking. In both CobJ and CbiE methylation occurs along the west edge of the docked ligand. Again, this could

be confirmed as suggested in chapter 4 by mixing purified PC6B with separate methyltransferase enzyme solutions and the use of mass spectrometry to confirm if an additional methylation has occurred.

It was surprising to observe that corrin ring decoration sequences did not dictate specificity through variation in the length of interacting amino acid R-groups. The amino acid interactions from each docking study are summarised together in Table 27. It might have been the case that longer amino acid R-groups would be seen interacting with acetic acid chains, whereas shorter R-groups would be seen interacting with propanoic acid chains. This trend was not observed so it cannot be concluded if order of decoration helps drive specificity upon binding. It is possible that perhaps the order and decoration of methylations also assists through London-dispersion forces/hydrophobic interactions, however these would be small and unlikely to drive the docking[116]. It can be concluded that the majority of interactions do come about via serine and arginine residues but other amino acids and even the protein backbone can assist in docking. CobJ is the only methyltransferase that interacts with the C1 ethanone decoration, which faces inwardly towards the active site, whereas when a precorrin ligand is bound in a clockwise fashion this faces outwardly as observed in CobM and CobF. It would be interesting to probe this decoration further in CobA and CobI, once structures are obtained, as this is a more unique decoration component of the corrin ring in the first half of this part of the cobalamin biosynthesis pathway.

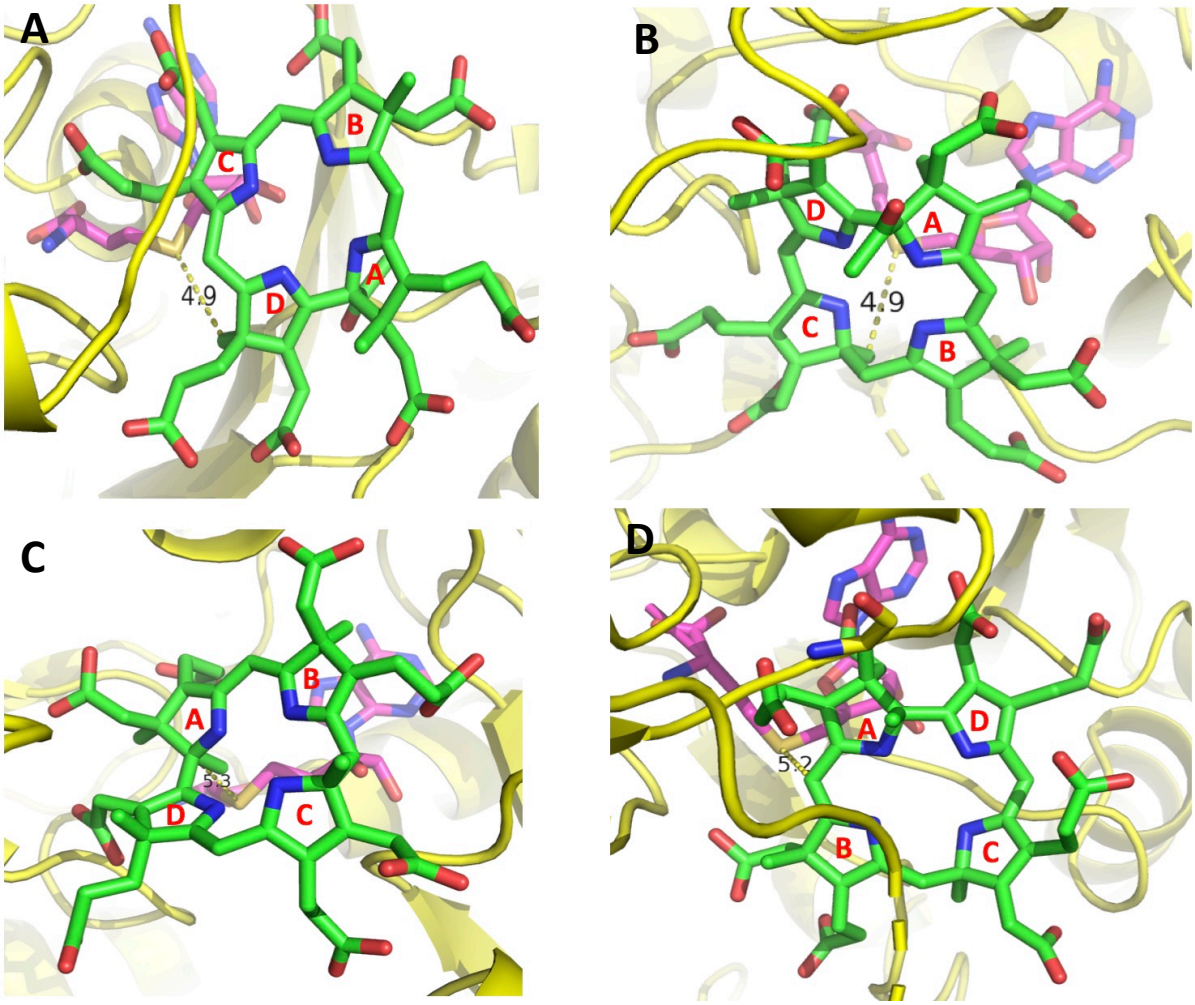


Figure 82: A summary of precorrin ligand docking to show clockwise and anticlockwise orientation where the pyrrole rings have been labelled. (A) PC4 to CobJ (3NUT.pdb) (B) PC4 to CobM (3NDC.pdb) (C) PC6A to CobF (3ND1.pdb) (D) PC6B to CbiE (2BB3.pdb). SAH is coloured purple, precorrin intermediates are coloured green and the methyltransferase protein is coloured yellow. Figure made via PyMOL, the PyMOL Molecular Grapics System, Version 2.0 Schrodinger, LLC.

Location of precorrin decoration	CobJ	CobM	CobF	CbiE
C1 (ethanone)	**Backbone** S137- 3.4 **R-group** S137- 3.0 N164-3.4	-	-	-
C2 (acetic acid)	**Backbone** S137- 3.4 N139-3.3 **R-group** N139-3.2	**Backbone** S35-3.0 L36-3.3	-	**R-group** S82-3.2 **Other** SAH-3.2
C3 (propanoic acid)	**R-group** S169-3.0	**R-group** T219-2.9	-	**Backbone** G42-3.1 L80-3.4 **R-group** H45-3.0
C7 (acetic acid)	-	-	**R-group** D39-1.8 S41-3.0	**Backbone** I90-3.3
C8 (propanoic acid)	-	**Backbone** L165-3.3 **R-group** R218-2.9	**Backbone** V242-3.3 **R-group** R232-3.0 R236-3.5	-
C12 (acetic acid)	**Backbone** N55-3.1 **R-group** N55-3.2 S80-3.0	**R-group** R137-3.1	**R-group** R169-3.1	**R-group** E152-3.4
C13 (propanoic acid)	**Backbone** S80-3.2 **R-group** S79-3.2 S80-3.0	-	**Backbone** L187-3.4 **R-group** R169-3.3	**R-group** E60-3.4
C17 (propanoic acid)	**R-group** R241-3.1	**Backbone** S81-3.1	**R-group** S120-2.8	**Backbone** L202-3.2 **R-group** R44-2.9 S199-3.4
C18 (acetic acid)	-	**R-group** T54-2.9	**Backbone** S120-3.0 **R-group** S120-2.9	**R-group** R44-2.9

Table 27: A summary of all the interacting amino acid residues from the docking studies in section 5.2.

5.5.3 Exploring the sequences of the β-loops and precorrin ligand decoration patterns

By exploring the sequences of loop regions of the tertiary structure involved in precorrin ligand docking, it was thought that the binding interactions could be elucidated further. Given that the precorrin decoration does not appear to drive specificity perhaps loop regions interact with a certain geometry of the docked precorrin ligand. This only appeared to be the case in the post β7 loop, which always interacted with the south-east pyrrole ring. None of the other loop regions showed complete unity across the methyltransferase discussed. However, it could be concluded that the post β2, β3 and β4 loop were involved with interactions to the west side of docked precorrin ligands, whereas the post β6, β7 and post β10 loop interacted with the east side of precorrin ligands. Given the architecture of the canonical methyltransferases and similarity in SAM/SAH binding orientation this is not surprising, given the location of methylation typically occurred to the west side with SAM buried underneath the substrate.

When these loop sequences were probed further to see if any similarities were present across these enzymes only the conserved motif GDX in the post β4 loop was consistent. It is interesting to note that the interacting serine residue in this loop is normally before the GDX motif, except in the case of CobF, which could be accountable to the ancillary reaction which has to occur before methylation can take place. The post β6 loop appeared to always contain at least one basic residue, typically an arginine and a serine, which again is not surprising given most of the precorrin docking interactions occur predominantly via these residues. These loops varied in length and it appears docking is a result of a contribution of interactions rather than one or two key residues.

5.5.4 Concluding remarks

These results show that the aerobic and anaerobic pairs are remarkably similar genetically and structurally, except in the case of CobF and CbiD. This is due to a varying ring contraction mechanism carried out by the aerobic pathway and completed by the ancillary reaction of CobJ. These docking studies show how these enzymes bind the precorrin ligands to be modified by SAM in the correct

orientation and geometry for methyl transfer. Some appear to do this without too firm a grip on the substrate, such as CobM. Others however have an apparently firmer grip such as CobF and CbiE which appear to have conserved basic residues that melt on to the substrate. CobJ was the only methyltransferase that required use of a normal mode to open the substrate-binding cleft slightly to allow product binding. There are lots of basic residues around to bind the substrate in CobJ, but surprisingly these are not conserved across the CobJ family suggesting maintenance of the general electrostatic potential is conserved rather than specific residues. The strength of these interactions is in contrast to those recently described for the product binding enzymes CobK and CobH which have several conserved basic residues that interact with the substrate carboxylates and clamp firmly around substrate and further embrace the product.

In conclusion there are no apparent conserved amino acid residues present in each methyltransferase to dictate specificity or, that the length of amino acids may play a role in matching to specific decoration components of the precorrin ligands. Exploration of the tertiary fold of the canonical methyltransferases was able to narrow down the post β-strand loop regions that play a role in docking. There did not appear to be any key defining steric features or similarities across these loop regions.

Furthermore, the methyltransferases of cobalamin biosynthesis may have similar architectures, with the exception of the C-terminal domain of CbiET/CobL, but the pattern and characteristics of substrate binding form a varied patchwork of interactions within the conserved framework and are able to bring about the essential methylations key to producing cobalamin. These specific loop regions have been identified and would provide an excellent target site for the creation of new cobalamin biomolecules, either through permitting new corrin ring decorations to dock or assist in modifying these decorations.

Chapter 6: Conclusions and Future Work

6.1 Concluding remarks of crystallography

To explore the product binding activity of CobK and CobH, the structures of the subsequent enzymes of the cobalamin biosynthetic pathway, CobL and CobB needed to be solved. These structures may have indicated regions for potential protein-protein interactions involved in the release of their substrates. Furthermore, the precise way by which these possible interactions are facilitated could have been probed further using site directed mutagenesis.

However, the full-length structures of CobL and CobB could not be solved, even though crystals were successfully obtained which resulted in diffraction data being collected. Given that over many years numerous structures of the biosynthetic enzymes have been solved in the cobalamin pathway, it is of interest to consider why CobL and CobB have not. It is clear from the present work that these are not as readily crystallised so as to generate meaningful data. Therefore, one reason that they are some of the last cobalamin biosynthetic enzymes to still not have had their structures solved is that other research groups may also have tried and also been unsuccessful. This log jam will be resolved by finding the enzymes from a species that yields high quality crystals.

CobL had the advantage that a structure of CbiET, the anaerobic equivalent, has been solved providing a very good model for molecular replacement. However, the quality and size of the CobL crystals obtained was not sufficient to attain an optimum resolution higher than three angstroms. Suitable models from the cobalamin biosynthesis pathway for CobB would have included the anaerobic equivalent CbiA, or CobQ, an enzyme further upstream, but neither of these have had their own structures solved. Despite confident models of the ATP binding domain and class I glutamine-dependent amidotransferase domain, both present in CobB, none of these seemed

sufficient during molecular replacement efforts. To tackle this anomalous dispersion using selenomethionine was attempted, but in a similar fashion to CobL the crystal quality and the quality of the diffraction data collected were not sufficient to solve the structure.

Co-crystallisation of the enzyme pairs CobK/L and CobH/B was somewhat prematurely attempted, as further experimental research should have been carried out to confirm if a protein-protein interaction may be occurring, indicating the likelihood of this being trapped in a crystal lattice. However, co-crystallisation of the enzyme pair CobK/L did result in a higher resolution structure of the apo-form of CobK, in a different space group, possible due to a transient interaction between CobL and CobK resulting in different protein contacts presenting in the packing and growing of the crystal. The structures solved of CobK and CobH did not reveal anything new when compared to existing structures. However, the flexible active site loop region of CobK was able to be modelled in the apoenzyme, due to crystal contacts present, highlighting how flexible this region is. An interesting avenue observed but not pursued due to time constraints of the project did elude to the possibility of an interaction between CobL, CobH and CobB. This was when the HBA signal recorded during thermal melt studies was not present when all these enzymes had the suitable substrates to facilitate the reaction pathway of PC6B to HBAD, in which the HBA signal was lost possibly because it was buried and hidden within a protein complex.

6.2 Concluding remarks of CobK product trapping activity and release

The product binding activity of CobK was confidently demonstrated, when in the presence of NADPH, through the use of experimental techniques including UV-VIS spectroscopy, ITC and NMR. This is in agreement with the existing PC6B purification protocols which rely on the isolation of product bound protein, which is subsequently removed after nickel column chromatography. The product bound crystal structure of CobK infers tight binding and encapsulation of PC6B however it is still not fully clear why CobK may do this. Given the similarity of PC6A and PC6B this could simply be

a case of mistaken identity however kinetics data and NMR showed that in the presence of CobL and CobL^{c} respectively PC6B is quickly released. Given the orientation of PC6B in the active site of CobK it suggests that it is presented to CobL with the C15 methylation readily available. This advocates a protein-protein interaction is more likely to be occurring, but this may only be a transient contact brought about by encouraging the opening of the active site $\beta2\beta3$ loop rather than full protein docking.

A possible argument for product trapping activity is to protect intermediates which may be labile or be lost to off pathway reactions. Given the similarity of the cobalamin intermediates subsequent enzymes in the pathway may be able to act out of turn, which would impact on the overall cobalamin synthesis. For example, PC6B may also be able to bind to CobB, missing out CobL and CobH, or alternatively HBA may be able to bind to CobQ, which carries out an identical chemical modification to CobB to different locations of the corrin ring. Therefore, product binding may ensure safe delivery of the substrate to the correct enzyme within the biosynthetic pathway of cobalamin.

This has been highlighted from further CobK studies showing that it does not only bind its substrate PC6A and product PC6B, but also the incorrectly C5-methylated PC6B, observed in a crystal structure, showing that a previous methyltransferase has acted out of turn, due to a case of mistaken identity. This and preliminary observations using NMR and fluorescence spectroscopy demonstrate that HBA can also incorrectly bind to CobK. This certainly adds weight to the proposal that these enzymes exhibit product binding activity to protect intermediates from other enzymes which includes those within the biosynthetic pathway itself. The protein-protein interaction shown between CobK and CobL would ensure the safe delivery of PC6B and CobH may also interact with CobL to ensure HBA is delivered to CobB. Therefore, the techniques used to probe interactions between CobK and CobL could be extended to CobL and CobH, CobH and CobB, and possibly a larger multimeric protein complex between three or all four of these enzymes.

6.3 Concluding remarks of canonical methyltransferases

Finally, the *in silico* docking studies of the canonical methyltransferases contrasts with the stronger binding affinities described for the product binding enzymes CobK and CobH, which contain several conserved basic residues to clamp and encapsulate their substrate/product. The specificity of the canonical methyltransferases, which have a highly conserved protein architecture, but are able to differentiate between the various precorrin intermediates has been explored. However it is still not clear how such specificity is brought about to differentiate these, given the lack of any conserved residues which contribute to the differentiation between these highly similar intermediates. The binding orientation of the subsequent pattern of corrin ring decoration also did not show a clear way to distinguish between these precorrin intermediates. Finally, the loop regions of these proteins were investigated further to determine whether the R-groups were sterically able to detect and dictate different precorrin docking orientations. However, it was not possible to conclusively show any defining features able to dictate such substrate specificity.

6.4 Future work

Regarding crystallography it is encouraging crystals can be obtained, however further work needs to be undertaken to optimise the crystal quality of the unsolved enzymes. Heavy metal labelling using mercury would increase the signal of anomalous diffraction data and since CobB contains an active site cysteine mercury would therefore be an attractive option. Given the time and effort already invested in this project, it may be more beneficial to attempt the crystallisation of CobB and CobL from different gene sequences, as there is a chance this will lead to crystals more readily and which are of better quality.

A lot more time could be spent refining some of the techniques used in this project such as ITC and NMR to continue to probe protein-protein interactions and incorrect substrate binding. Enzyme assays to explore cases of mistaken identity for various Cob enzymes would include glutamine or ATP assays to track CobB activity and SAH assays to track the activity of the methyltransferases.

A further Cob protein that may have been overlooked is CobE. The synthesis of cobalamin does proceed without the need of this enzyme, however Deery et al, 2012[53] provides comprehensive evidence to support that CobE is a chaperone for various cobalamin intermediates including PC4, PC5, PC6B, PC8 and HBA. It has had its crystal structure solved (2W6K.pdb and 2W6L.pdb), but it is surprising that it has not been crystallised with any of these intermediates. Perhaps co-crystallisation studies including CobE would provide a better insight into the co-ordination of cobalamin biosynthetic enzymes, offering an additional but insightful layer to the complexity of the pathway, especially if a multimeric protein complex exists between some of the Cob proteins.

The canonical methyltransferases play a large role in the production of cobalamin. Through docking analysis, loop regions important for substrate binding have been identified. These narrow down the options for mutation sites, which could be coupled with SAH enzyme assays, to further explore and identify residues important for dictating specificity. If successful, this would provide the grounding to help facilitate the design and synthesis of new cobalamin molecules, where further mutagenesis studies could facilitate the accommodation or assist in addition of new components to these desired cobalamin-based molecules.

References

1. Raux, E., Schubert, H. L. & Warren, M. J. Biosynthesis of cobalamin (vitamin B 12): A bacterial conundrum. *Cell. Mol. Life Sci.* **57,** 1880–1893 (2000).

2. Warren, M. J., Raux, E., Schubert, H. L. & Escalante-Semerena, J. C. The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* **19,** 390–412 (2002).

3. Subramanian, G., Zhang, X., Kodis, G, Kong, Q., Liu, C., Chizmeshya, A., Weierstall, U. & Spence, J. Direct structural and chemical characterization of the photolytic intermediates of methylcobalamin using time-resolved X-ray absorption spectroscopy. *J. Phys. Chem. Lett.* **9,** 1542–1546 (2018).

4. Bridwell-Rabb, J. & Drennan, C. L. Vitamin B12 in the spotlight again. *Curr. Opin. Chem. Biol.* **37,** 63–70 (2017).

5. Warren, M. J. & Escalante-Semerena, J. C. Biosynthesis and use of cobalamin (B12). *EcoSal Plus* **1,** (2008).

6. Froese, D. S., Fowler, B. & Baumgartner, M. R. Vitamin B12 , folate, and the methionine remethylation cycle-biochemistry, pathways, and regulation. *J. Inherit. Metab. Dis.* **42,** 1–13 (2019).

7. Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. The role of B vitamins in marine biogeochemistry. *Ann. Rev. Mar. Sci.* **6,** 339–367 (2014).

8. Webb, M. E., Marquet, A., Mendel, R. R., Rébeillé, F. & Smith, A. G. Elucidating biosynthetic pathways for vitamins and cofactors. *Nat. Prod. Rep.* **24,** 988 (2007).

9. Randaccio, L., Geremia, S. & Wuerges, J. Crystallography of vitamin B12 proteins. *J. Organomet. Chem.* **692,** 1198–1215 (2007).

10. Okuda, K. Discovery of vitamin B12 in the liver and its absorption factor in the stomach: A historical review. *J. Gastroenterol. Hepatol.* **14,** 301–308 (1999).

11. Battersby, A. How nature builds the pigments of life: the conquest of vitamin B12. *Science* **264,** 1551–1557 (1994).

12. Rickms, E. L., Brink, N. G., Koniuszy, F. R., Wood, T. R. & Folkers, K. Crystalline vitamin B12. *Science* **107,** 396–397 (1948).

13. Hodgkin, D. C., Kamper, J., Lindsey, J., Maureen, M., Pickworth, J. & Robertson, J. H.. The structure of vitamin B12 . I. An outline of the crystallographic investigation of vitamin B12. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **242,** 228–263 (1957).

14. Roth, J. R., Lawrence, J. G. & Bobik, T. A. Cobalamin (coenzyme B12): synthesis and biological significance. *Annu. Rev. Microbiol.* **50,** 137–181 (1996).

15. Raux, E., Schubert, H. L., Roper, J. M., Wilson, K. S. & Warren, M. J. Vitamin B12 : Insights into biosynthesis's mount improbable. *Bioorg. Chem.* **27,** 100–118 (1999).

16. Degnan, P. H., Taga, M. E. & Goodman, A. L. Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab.* **20,** 769–778 (2014).

17. MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J. & Walhout, A. J. M. Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell* **153,** 240–252 (2013).

18. Watson, E., MacNeil, L., Ritter, A., Yilmaz, L., Rosebrock, A., Caudy, A. & Walhout, A. Interspecies systems biology uncovers metabolis affecting *C. elegans* gene expresion and life history traits. *Cell* **156,** 759–770 (2014).

19. Taga, M. E. & Walker, G. C. Sinorhizobium meliloti requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. *Mol. Plant-Microbe Interact.* **23,** 1643–1654 (2010).

20. Brouwer-Brolsma, E. M., Dhonukshe-Rutten, R. A.M., van Wijngaarden, J.P., van der Zwaluw, N. L., van der Velde, N. & de Groot, L. C.P.G.M. Dietary sources of vitamin B-12 and their association with vitamin B-12 status markers in healthy older adults in the B-PROOF study. *Nutrients* **7,** 7781–7797 (2015).

21. Green, R., Allen, L. H., Bjorke-Monsen, A., Brito, A., Gueant, J., Miller, J., Molloy, A., Nexo, E., Stabler, S., Toh, B., Ueland, P. M. & Yanjnik, C. Vitamin B12 deficiency. *Nat. Rev. Dis. Prim.* **3,** 1–20 (2017).

22. Yajnik, C. S., Deshpande, S S., Lubree, H. G., Naik, S. S., Bhat, D. S., Uradey, B. S., Deshpande, J. A., Rege, S. S., Refsum, H. & Yudkin, J. S. Vitamin B12 deficiency and hyperhomocysteinemia in rural and urban indians. *J. Assoc. Physicians India* **54,** 775–782 (2006).

23. Lu, H., Liu, X., Deng, Y. & Qing, H. DNA methylation, a hand behind neurodegenerative diseases. *Front. Aging Neurosci.* **5,** 1–16 (2013).

24. Banihani, S. A. Vitamin B12 and semen quality. *Biomolecules* **7,** 1–10 (2017).

25. Bottiglieri, T., Laundy, M., Crellin, R., Toone, B. K., Carney, M. W. P. & Reynolds, E. H. Homocysteine, folate, methylation, and monoamine metabolism in depression. *J Neurol Neurosurg Psychiatry* **69,** 228–232 (2000).

26. Banerjee, R. & Ragsdale, S. W. The many faces of vitamin B12: Catalysis by cobalamin-dependent enzymes. *Annu. Rev. Biochem.* **72,** 209–247 (2003).

27. Matthews, R. G. Cobalamin-dependent methyltransferases. *Acc. Chem. Res.* **34,** 681–689 (2001).

28. Yao, L., Yu, L., Zhang, J., Xie, X., Tao, Q., Yan, X., Hong, Q., Qiu, J., He, J. & Ding, D. A tetrahydrofolate-dependent methyltransferase catalyzing the demethylation of Dicamba in *Sphingomonas* sp. strain Ndbn-20. *Appl. Environ. Microbiol.* **82,** 5621–5630 (2016).

29. Marsh, E. N. G. & Drennan, C. L. Adenosylcobalamin-dependent isomerases: New insights into structure and mechanism. *Curr. Opin. Chem. Biol.* **5,** 499–505 (2001).

30. Banerjee, R. The Yin-Yang of cobalamin biochemistry. *Chem. Biol.* **4,** 175–186 (1997).

31. Shelton, D. & Tiedje, J. Isolation and partial characterization of bacteria in anaerobic consortium that mineralizes 3-chlorobenzoic acid. *Appl. Environ. Microbiol.* **48,** 394–400 (1984).

32. Tian, H., Shimakoshi, H., Park, G., Kim, S., You, Y. & Hisaeda, Y. Photocatalytic function of the B12 complex with the cyclometalated iridium (III) complex as a photosensitizer under visible light irradiation. *Dalt. Trans.* **47,** (2018).

33. Smidt, H., Van Leest, M., Van der Oost, J. & De Vos, W. M. Transcriptional regulation of the cpr gene cluster in ortho-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *J. Bacteriol.* **182,** 5683–5691 (2000).

34. Christiansen, N., Ahring, B. K., Wohlfarth, G. & Diekert, G. Purification and characterization of the 3-chloro-4-hydroxy-phenylacetate reductive dehalogenase of *Desulfitobacterium hafniense*. *FEBS Lett.* **436,** 159–162 (1998).

35. de Vos, W. M. *et al.* Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *J. Biol. Chem.* **274,** 20287–20292 (2002).

36. Lawrence, A. D., Nemoto-Smith, E., Deery, E., Baker, J. A., Schroeder, S., Brown, D. G., Tullet, J. M. A., Howard, M. J., Brown, I. R., Smith, A. G., Boshoff, H. I., Barry, C. E. & Warren, M. J . Construction of fluorescent analogs to follow the uptake and distribution of cobalamin (vitamin B12) in bacteria, worms, and plants. *Cell Chem. Biol.* **25,** 1–11 (2018).

37. Waibel, R., Treichler, H., Schaefer, N. G., van Staveren, D. R., Mundwiler, S., Kunze, S., Kuenzi, M., Alberto, R., Nuesch, J., Knuth, A., Moch, H., Schibli, R., & Schubiger, P. A . New derivatives of vitamin B12 show preferential targeting of tumors. *Cancer Res.* **68,** 2904–2911 (2008).

38. Collins, D. A. Imaging cobalamin uptake within malignant breast tumors *in vivo*. *Mol. Imaging Biol.* **21,** 1–12 (2018).

39. Flodh, H. & Ullberg, S. Accumulation of labeled vitamin B12 in some transplanted tumours. *Int. J. Cancer* **3,** 694–699 (1968).

40. Birn, H., Nexø, E., Christensen, E. I. & Nielsen, R. Diversity in rat tissue accumulation of vitamin B12 supports a distinct role for the kidney in vitamin B12 homeostasis. *Nephrol. Dial. Transplant.* **18,** 1095–1100 (2003).

41. Lodowski, P., Toda, M. J., Ciura, K., Jaworska, M. & Kozlowski, P. M. Photolytic properties of antivitamins B12. *Inorg. Chem.* **57,** 7838–7850 (2018).

42. Rury, A. S., Wiley, T. E. & Sension, R. J. Energy cascades, excited state dynamics, and photochemistry in cob(III)alamins and ferric porphyrins. *Acc. Chem. Res.* **48,** 860–867 (2015).

43. Shell, T. A. & Lawrence, D. S. Vitamin B12: A tunable, long wavelength, light-responsive platform for launching therapeutic agents. *Acc. Chem. Res.* **48,** 2866–2874 (2015).

44. Rodgers, Z. L., Hughes, R. M., Doherty, L. M., Shell, J. R., Molesky, B. P., Brugh, A. M., Forbes, M. D. E., Moran, A. M. & Lawrence, D. S. B12-mediated, long wavelength photopolymerization of hydrogels. *J. Am. Chem. Soc.* **137,** 3372–3378 (2015).

45. Scott, A. I. & Roessner, C. A. Biosynthesis of cobalamin (vitamin B12). *Biochem. Soc. Trans.* **30,** 613–620 (2002).

46. Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. & Thermes, C. *Salmonella typhimurium* cobalamin (vitamin B12) biosynthetic genes: Functional studies in *S. typhimurium* and *Escherichia coli*. *J. Bacteriol.* **178,** 753–67 (1996).

47. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B12: A review and future perspectives. *Microb. Cell Fact.* **16,** 15 (2017).

48. Men, Y., Seth, E. C., Yi, S., Allen, R. H., Taga, M. E. & Alvarez-Cohen, L. Sustainable growth of *Dehalococcoides mccartyi* 195 by corrinoid salvaging and remodeling in defined lactate-fermenting consortia. *Appl. Environ. Microbiol.* **80,** 2133–2141 (2014).

49. Rosnow, J. J., Hwang, S., Killinger, B. J., Kim, Y., Moore, R. J., Lindemann, S. R., Maupin-Furlow, J. A. & Wright, A. T. Cobalamin activity-based probe enables microbial cell growth and finds new cobalamin-protein interactions across domains. *Appl. Environ. Microbiol.* **84,** 1 (2018).

50. Uzar, H. C., Battersby, A., Carpenter, A. & Leeper, F. J. Biosynthesis of porphyrins and related macrocycles. Development of a pulse labelling method to determine the C-methylation sequence for vitamin B12. *J. Chem. Soc.* **11,** (1987).

51. Debuscche, L., Thibaut, D., Cameron, B., Crouzet, J. & Blanche, F. Purification and characterization of cobyrinic acid a,c-diamide synthase from *Pseudomonas denitrificans*. *J. Bacteriol.* **172,** 6239–6244 (1990).

52. Scott, A. I. Discovering nature's diverse pathways to vitamin B12: A 35-year odyssey. *J Org Chem* **68,** 2529–2539 (2003).

53. Deery, E., Schroeder, S., Lawrence, A. D., Taylor, S. L., Seyedarabi, A., Waterman, J., Wilson, K. S., Brown, D. G., Geeves, M. A., Howard, M. J., Pickersgill, R. W. & Warren, M. J. An enzyme-trap approach allows isolation of intermediates in cobalamin biosynthesis. *Nat. Chem. Biol.* **8,** 933–940 (2012).

54. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: A chronicle of convergence. *Trends Biochem. Sci.* **28,** 329–335 (2003).

55. Warren, M. J., Stolowich, N. J., Santander, P. J., Roessner, C. A., Sowa, B. A. & Scott, I. A. Enzymatic synthesis of dihydrosirohydrochlorin (precorrin-2) and of a novel pyrrocorphin by uroporphyrinogen III methylase. *FEBS Lett.* **261,** 76–80 (1990).

56. Warren, M. J., Roessner, C. A., Ozaki, S., Stolowich, N. J., Santander, P. J. & Scott, A. I. Enzymatic synthesis and structure of precorrin-3, a trimethyldipyrrocorphin intermediate in vitamin B12 biosynthesis. *Biochemistry* **31,** 603–609 (1992).

57. Scott, I. A., Roessner, C. A., Stolowich, N. J., Spencer, J. B., Min, C. & Ozaki, S. I. Biosynthesis of vitamin B12. Discovery of the enzymes for oxidative ring contraction and insertion of the fourth methyl group. *FEBS Lett.* **331,** 105–108 (1993).

58. Min, C., Atshaves, B. P, Roessner, C. A., Stolowich, N. J., Spencer, J. B. & Scott, A. I. Isolation, structure, and genetically engineered synthesis of precorrin-5, the pentamethylated intermediate of vitamin B12 biosynthesis. *J Am Chem Soc* **115,** 10380–10381 (1993).

59. Thibaut, D., Debussche, L. & Blanche, F. Biosynthesis of vitamin B12: Isolation of precorrin-6x, a metal-free precursor of the corrin macrocycle retaining five S-adenosylmethionine-derived peripheral methyl groups. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8795–8799 (1990).

60. Roessner, C. A., Spencer, J. B., Stolowich, N. J., Wang, J., Nayar, G. P., Santander, P. J., Pichon, C., Min, C., Holderman, M. T. & Scott, A. I. Genetically engineered synthesis of precorrin-6x and the complete corrinoid, hydrogenobyrinic acid, an advanced precursor of vitamin B12. *Chem. Biol.* **1,** 119 (1994).

61. Blanche, F., Thibaut, D., Famechon, A., Debussche, L., Cameron, B. & Crouzet, J. Precorrin-6x reductase from *Pseudomonas denitrificans*: Purification and characterization of the enzyme and identification of the structural gene. *J. Bacteriol.* **174,** 1036–1042 (1992).

62. Gu, S., Sushko, O., Deery, E., Warren, M. J. & Pickersgill, R. W. Crystal structure of CobK reveals strand-swapping between Rossmann-fold domains and molecular basis of the reduced precorrin product trap. *Sci. Rep.* **5,** 16943 (2015).

63. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250,** 194–199 (1974).

64. Blanche, F., Famechon, A., Thibaut, D., Debussche, L., Cameron, B. & Crouzet, J. Biosynthesis of vitamin B12 in *Pseudomonas denitrificans*: the biosynthetic sequence from precorrin-6y to precorrin-8x is catalyzed by the cobL gene product. *J Bacteriol* **174,** 1050–1052 (1992).

65. Keller, J. P., Smith, P. M., Benach, J., Christendat, D., DeTitta, G. T. & Hunt, J. F. The crystal structure of MT0146/CbiT suggests that the putative precorrin-8w decarboxylase is a methyltransferase. *Structure* **10,** 1475–87 (2002).

66. Thibaut, D., Couder, M., Famechon, A., Debussche, L., Cameron, B., Cameron, B. & Blanche, F. The final step in the biosynthesis of hydrogenobyrinic acid is catalyzed by the cobH gene product with precorrin-8x as the substrate. *J. Bacteriol.* **174,** 1043–1049 (1992).

67. Shipman, L. W., Li, D., Roessner, C. A., Scott, A. I. & Sacchettini, J. C. Crystal structure of precorrin-8x methyl mutase. *Structure* **9,** 587–96 (2001).

68. Galperin, M. Y. & Grishin, N. V. The synthetase domains of cobalamin biosynthesis amidotransferases CobB and CobQ belong to a new family of ATP-dependent amidoligases, related to dethiobiotin synthetase. *Proteins* **41,** 238–47 (2000).

69. Curti, B., Vanoni, M. A., Verzotti, E. & Zanetti, G. Glutamate synthase: A complex iron-sulphur flavoprotein. *Biochem. Soc. Trans.* **24,** 95–99 (1995).

70. Massière, F. & Badet-Denisot, M. A. The mechanism of glutamine-dependent amidotransferases. *Cell. Mol. Life Sci.* **54,** 205–222 (1998).

71. Van Den Heuvel, R. H. H., Curti, B., Vanoni, M. A. & Mattevi, A. Glutamate synthase: A fascinating pathway from L-glutamine to L-glutamate. *Cell. Mol. Life Sci.* **61,** 669–681 (2004).

72. Crampton, D. J., Lobrutto, R. & Frasch, W. D. Identification of the P-loop lysine as a metal ligand in the absence of nucleotide at catalytic site 3 of chloroplast F1-ATPase from *Chlamydomonas reinhardtii*. *Biochemistry* **40,** 3710–3716 (2001).

73. Leipe, D. D., Wolf, Y. I., Koonin, E. V & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317,** 41–72 (2002).

74. Walker, J. E., Saraste, M., Runswick, M. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951 (1982).

75. Tanwar, A. S., Morar, M., Panjikar, S. & Anand, R. Formylglycinamide ribonucleotide amidotransferase from *Salmonella typhimurium*: Role of ATP complexation and the glutaminase domain in catalytic coupling. *Acta Cryst* **D68,** 627–636 (2012).

76. List, F., Vega, M. C., Razeto, A., Häger, M. C., Sterner, R. & Wilmanns, M. Catalysis uncoupling in a glutamine amidotransferase bienzyme by unblocking the glutaminase active site. *Chem. Biol.* **19,** 1589–1599 (2012).

77. Tanwar, A. S., Sindhikara, D. J., Hirata, F. & Anand, R. Determination of the formylglycinamide ribonucleotide amidotransferase ammonia pathway by combining 3D-RISM theory with experiment. *ACS Chem. Biol.* **10,** 698–704 (2015).

78. Fresquet, V., Williams, L. & Raushel, F. M. Mechanism of cobyrinic acid a,c-diamide synthetase from *Salmonella typhimurium* LT2. *Biochemistry* **43,** 10619–10627 (2004).

79. Raushel, F. M., Thoden, J. B. & Holden, H. M. Enzymes with molecular tunnels. *Acc. Chem. Res.* **36,** 539–548 (2003).

80. Huang, X., Holden, H. M. & Raushel, F. M. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu. Rev. Biochem.* **70,** 149–80 (2001).

81. Milani, M., Psece, A., Bolognesi, M., Bocedi, A. & Ascenzi, P. Substrate channeling. *Biochem. Mol. Biol. Educ.* **31,** 289–297 (2003).

82. Lynch, E. M., Hicks, D. R., Shepherd, M., Endrizzi, J. A., Maker, A., Hansen, J. M., Barry, R. M., Gitai, Z., Baldwin, E. P. & Kollman, J. M. Human CTP synthase filament structure reveals the active enzyme conformation. *Nat. Struct. Mol. Biol.* **24,** 1–10 (2017).

83. Yang, G., Sandalova, T., Lohman, K., Lindqvist, Y. & Rendina, A. R. Active site mutants of *Escherichia coli* dethiobiotin synthetase: Effects of mutations on enzyme catalytic and structural properties. *Biochemistry* **36,** 4751–4760 (1997).

84. Dey, S., Lane, J. M., Lee, R. E., Rubin, E. J. & Sacchettini, J. C. Structural characterization of the *Mycobacterium tuberculosis* biotin biosynthesis enzymes 7,8-diaminopelargonic acid synthase and dethiobiotin synthetase. *Biochemistry* **49,** 6746–6760 (2010).

85. Taylor, G. L. Introduction to phasing. *Acta Cryst* **D66,** 325–338 (2010).

86. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Cryst* **F70,** 2–20 (2014).

87. McPherson, A. Introduction to protein crystallization. *Methods* **34,** 254–265 (2004).

88. Lamb, A. L., Kappock, T. J. & Silvaggi, N. R. You are lost without a map: Navigating the sea of protein structures. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1854,** 258–268 (2015).

89. Dauter, Z. & Jaskolski, M. How to read (and understand) Volume A of International Tables for Crystallography: An introduction for nonspecialists. *J. Appl. Crystallogr.* **43,** 1150–1171 (2010).

90. Taylor, G. The phase problem. *Acta Cryst* **D59,** 1881–1890 (2003).

91. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Cryst* **D67,** 235–242 (2011).

92. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R, Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Cryst* **D66,** 213–221 (2010).

93. Ealick, S. E. Advances in multiple wavelength anomalous diffraction crystallography. *Curr. Opin. Chem. Biol.* **4,** 495–499 (2000).

94. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst* **D66,** 486–501 (2010).

95. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D66,** 12–21 (2010).

96. Kleywegt, G. J. On vital aid: The why, what and how of validation. *Acta Cryst* **D65,** 134–139 (2009).

97. Dauter, Z., Lamzin, V. S. & Wilson, K. S. Proteins at atomic resolution. *Curr. Opin. Struct. Biol.* **5,** 784–790 (1995).

98. Dauter, Z. Current state and prospects of macromolecular crystallography. *Acta Cryst* **D62,** 1–11 (2006).

99. Bracken, C., Marley, J. & Lu, M. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20,** 71–75 (2001).

100. Trott, O. & Olson, A. J. AutoDock: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **28,** 455–461 (2009).

101. Suhre, K. & Sanejouand, Y. H. ElNémo: A normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32,** 610–614 (2004).

102. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4,** 17 (2012).

103. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34,** 130–135 (2001).

104. Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D. & Race, P. R. Improving the success rate of protein crystallization by random microseed matrix screening. *J. Vis. Exp.* 1–6 (2013).

105. Kleywegt, G. J. Validation of protein crystal structures. *Acta Cryst* **D56,** 249–265 (2000).

106. Kang, J., Kuroyanagi, S., Akisada, T., Hagiwara, Y. & Tateno, M. Unidirectional mechanistic valved mechanisms for ammonia transport in GatCAB. *J. Chem. Theory Comput.* **8,** 649–660 (2012).

107. Zhao, L., Rathnayake, U. M., Dewage, S. W., Wood, W. N., Veltri, A. J., Veltri, A. J. & Hendrickson, T. L. Characterization of tunnel mutants reveals a catalytic step in ammonia delivery by an aminoacyl-tRNA amidotransferase. *FEBS Lett.* **590,** 3122–3132 (2016).

108. Amaro, R. E., Sethi, A., Myers, R. S., Davisson, V. J. & Luthey-Schulten, Z. A. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* **46,** 2156–2173 (2007).

109. Endrizzi, J., Kim, H., Anderson, P. & Baldwin, E. Mechanisms of product feedback regulation and drug resistance in cytidine triphosphate synthetases from the structure of a CTP-inhibited complex. *Natl. Institutes Heal.* **15,** 1–18 (2005).

110. Miles, B. W., Thoden, J. B., Holden, H. M. & Raushel, F. M. Inactivation of the amidotransferase activity of carbamoyl phosphate synthetase by the antibiotic acivicin. *J. Biol. Chem.* **277,** 4368–4373 (2002).

111. Souza, J. O. De, Dawson, A. & Hunter, W. N. An improved model of the *Trypanosoma brucei* CTP synthetase glutaminase domain – acivicin complex. *ChemMedChem* **12,** 577–579 (2017).

112. Vuignier, K., Schappler, J., Veuthey, J. L., Carrupt, P. A. & Martel, S. Drug-protein binding: A critical review of analytical tools. *Anal. Bioanal. Chem.* **398,** 53–66 (2010).

113. Fang, H., Li, D., Kang, J., Jiang, P., Sun, J. & Zhang, D. Metabolic engineering of *Escherichia coli* for de novo biosynthesis of vitamin B12. *Nat. Commun.* **9,** 1–12 (2018).

114. Harris, G. & Scott, D. J. Isothermal titration calorimetry: Feeling the heat. *Biochem. (Lond).* **41,** 4–7 (2019).

115. Cornish-Bowden, A. Why is uncompetitive inhibition so rare? A possible explanation, with implications for the design of drugs and pesticides. *FEBS Lett.* **203,** 3–6 (1986).

116. Struck, A. W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-adenosyl-methionine-dependent methyltransferases: Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* **13,** 2642–2655 (2012).

117. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** 320–324 (2014).

118. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).